

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: September 21, 2004, 15:06:11 ; Search time 71 Seconds
(without alignments)

6541.772 Million cell updates/sec

Title: US-10-776-827-10

Perfect score: 8026

Sequence: 1 cggggcagcaaccaggagat.....aattaaaggatatactagctc 4460

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 283362

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=framer_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10776827/runat_21092004_133240_20767/app_query.fasta_1.4615
-DB=SwissProt 42 -QFMT=fastan -SUFFIX=n2p.rsp -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEARSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10776827@cgn 1 1 120 @runat_21092004_133240_20767 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THRBADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	850	10.6	153	1 Y040 HUMAN	Q15053 homo sapien
2	123.5	1.5	319	1 TCBI_RABIT	P06333 oryctolagus
3	117.5	1.5	628	1 V70K TYMV	P10357 turnip yell
4	113	1.4	628	1 V70K TYMVC	P28478 turnip yell
5	112.5	1.4	631	1 HNEFA HUMAN	P20823 homo sapien
6	112	1.4	597	1 HNEFA TYMV	P20130 ononis yell
7	108.5	1.4	1007	1 SAL2 HUMAN	Q9Y467 homo sapien
8	108	1.3	353	1 ALG1_GORGO	P20758 gorilla gor
9	108	1.3	862	1 Z409 HUMAN	Q9up66 homo sapien
10	107.5	1.3	2063	1 NC06 HUMAN	Q14686 h nuclear r
11	107	1.3	628	1 V70K TYMVA	P20131 turnip yell
12	106.5	1.3	1004	1 SAL2_MOUSE	Q9QX96 mus musculus
13	106.5	1.3	2843	1 APC_HUMAN	P25054 homo sapien
14	106	1.3	5179	1 MUC2_HUMAN	Q02817 homo sapien
15	105	1.3	1964	1 NT44_MOUSE	P31695 mus musculus
16	103	1.3	353	1 ALG1_HUMAN	P01876 homo sapien
17	103	1.3	917	1 SMOO_HUMAN	P53814 homo sapien
18	102.5	1.3	2842	1 APC_RAT	P70478 rattus norv

c	19	102	1.3	962	1	GUNA_PSEFL	P10476 pseudomonas
	20	101.5	1.3	628	1	HNEFA_RAT	P15257 rattus norv
c	21	101.5	1.3	769	1	ITE2_BOVIN	P32592 bos taurus
	22	101.5	1.3	1185	1	DRPL_HUMAN	P54259 homo sapien
	23	101	1.3	817	1	HUNB_MUSDO	Q01778 musca domes
	24	101	1.3	1959	1	AGRI_RAT	P25304 rattus norv
	25	101	1.3	2003	1	NTC4_HUMAN	Q99466 homo sapien
	26	100.5	1.3	703	1	ZMI5_HUMAN	Q9H091 homo sapien
	27	100.5	1.3	977	1	BAB1_DROME	Q9W0K7 drosophila
	28	100.5	1.3	1210	1	AF4_HUMAN	P51825 homo sapien
	29	100.5	1.3	1509	1	GSRI_HUMAN	Q9NZM4 homo sapien
	30	100	1.2	1257	1	PGCN_RAT	P55067 rattus norv
	31	99.5	1.2	648	1	FXN1_HUMAN	O15353 homo sapien
	32	99.5	1.2	660	1	YHL1_EBV	P03181 epstein-bar
	33	99.5	1.2	1229	1	P121_HUMAN	Q9Y2N3 homo sapien
	34	99	1.2	315	1	SK12_HUMAN	O15370 homo sapien
	35	99	1.2	759	1	HUNB_DROVA	O62541 drosophila
	36	99	1.2	2453	1	NCR1_MOUSE	Q60974 mus musculus
	37	98.5	1.2	2845	1	APC_MOUSE	Q61315 mus musculus
	38	98	1.2	1426	1	BCL9_HUMAN	O00512 homo sapien
	39	98	1.2	1845	1	Z236_HUMAN	Q9UL36 homo sapien
	40	97.5	1.2	448	1	EDAR_MOUSE	Q9R187 mus musculus
	41	97	1.2	1544	1	TUL4_HUMAN	Q9NRJ4 homo sapien
	42	96.5	1.2	344	1	CD2_RAT	P08921 rattus norv
	43	96.5	1.2	628	1	HNEFA_MOUSE	P22361 mus musculus
	44	96.5	1.2	693	1	CAUP_DROME	P54269 drosophila
	45	96.5	1.2	1183	1	DRPL_RAT	P54258 rattus norv

ALIGNMENTS

RESULT 1

Y040 HUMAN	153 AA
ID Y040 HUMAN	STANDARD; PRT; 153 AA.
AC Q15053;	
DT 15-JUL-1998 (Rel. 36, Created)	
DT 15-JUL-1998 (Rel. 36, Last sequence update)	
DT 16-OCT-2001 (Rel. 40, Last annotation update)	
DE Hypothetical protein KIAA0040.	
GN KIAA0040.	
OS Homo sapiens (Human).	
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
ON NCBI_TaxID=9606;	
OX [1]	
RP SEQUENCE FROM N.A.	
RC TISSUE=Bone marrow;	
RX MEDLINE=96051387; PubMed=7584026;	
RA Nomura N., Miyajima N., Sazuka T., Tanaka A., Kawarabayashi Y.,	
RA Sato S., Nagase T., Seki N., Ishikawa K.-I., Tabata S.;	
RT "Prediction of the coding sequences of unidentified human genes. I.	
RT The coding sequences of 40 new genes (KIAA0001-KIAA0040) deduced by	
RT analysis of randomly sampled cDNA clones from human immature myeloid	
RT cell line KG-1."	
RL DNA Res. 1:27-35(1994).	
RL [2]	
RN SEQUENCE FROM N.A.	
RA Bird C.;	
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.	

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; D25539; BAA05022.1; --
DR EMBL; Z99715; CAB16870.1; --
KW Hypothetical protein.
SQ SEQUENCE 153 AA; 17035 MW; CAC3F8AF16821935 CRC64;

Alignment Scores:					
Pred. No.: Score: Percent Similarity: Best Local Similarity: Query Match: DB:	1_93e-66 850.00 100.00% 100.00% 10.59% 1	Length: Matches: Conservative: Mismatches: Indels: Gaps:	153 153 0 0 0 0		
US-10-776-827--10 (1-4460) x Y040_HUMAN (1-153)					
Qy	816	ATGCACGTATTCTCCAGTTCACCAAGGAACAACCCTCAGACTAAG	875		
Dd	1	MethisTyValHisValHisArgValThrGlnProArgAsnLysProGlnThrLys	20		
Qy	876	TGTCGCCAGTGAGGGCGAGTCCCAGGGACCGGTGAGCAAATCTTTGGNATGCTCTGGCCA	935		
Dd	21	CysProSerGlyGlyGlnSerGingLyProArqgLyGlnPheLeuAspThrValLeuAla	40		
Qy	936	GCTATGTGTCMAATAGAAGTCCTCTTACTGCAGACCCAGGCATGCTCCCCACTGTC TC	995		
Dd	41	AlaMet CysProfileAlaMeLeuleuthrAlalaaspProGlyMetProTrnCysleu	60		
Qy	996	TGGCATACCCCACATGCAAAAGAACATAATTATCCATACATCTCAATATGGTTCCC	1055		
Dd	61	TrpHistrProHisAlaLysHisLysGluHisLeuSeriIleHisLeuAsnMetValpro	80		
Qy	1056	AAGTGTGTGCATGCGAGTAACACACACACACAAAATTCAGGTACGAGTAGTGGGC	1115		
Dd	81	LysCysValHisMetHisValThrHisthrHisThrAsnSerGlyserArgTy rVaigly	100		
Qy	1116	AAGTATATTCTGCTCATCAATGCTCATTTGGCTATGTACTTTNGCAGGGAAGTACATTA	1175		
Dd	101	LystyrileLeuLeulleLysTripsSerLeualaMettyrPheValGinGlySerThrLeu	120		
Qy	1176	TCTACAGTCACAAAAATGTCTCATGGGAAGCCITTGCCAGATTTCAGACACATATATACAA	1235		
Dd	121	SerthrValThrLysMetSerHisGlYlysAlaLeuProAspSerAspThry-Ileglin	140		
Qy	1236	TTTTCTAACACGAAGGCCCCCATACCAATCTATTCCA	1274		
Dd	141	PheProAnGInGInglyProHisHisThrProSertllePro	153		

```

RESULT 2
TCBI_RABBIT STANDARD; PRT; 319 AA.
ID TCBI_RABBIT
AC P06333;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell receptor beta chain ANA 11.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
CX NCBI_TaxID=9986;
RN [1]
SEQUENCE FROM N.A.
RP MEDLINE=85242712; PubMed=2989826;
RX Angelillo A.L., Lamoyi E., Bernstein K.E., Mage R.G.;
RA "Identification of genes for the constant region of rabbit T-cell
RT receptor beta chains.";
RT receptor beta chains.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:4498-4502(1985).
RN [2]
REVIEWS TO 79-83 AND 104-105.
RP
RA Mage R.G.;
RL Submitted (APR-1987) to the PIR data bank.
DR PIR; A94052; EWRBB.
DR HSSP; P01842; 7FAB.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003597; Ig_C1.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00407; Igc1; 1.
DR PROSITE; PS50835; IG LIKE; 1.
DR T-cell. Receptor; Immunoglobulin domain; Transmembrane; Glycoprotein.

```

FT	DOMAIN	142	273	C REGION.
FT	TRANSMEM	292	313	CYTOPLASMIC TAIL.
FT	DOMAIN	314	319	N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT	CARBOHYD	147	147	N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT	CARBOHYD	164	164	N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT	CARBOHYD	254	254	N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT	CARBOHYD	262	282	N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT	CARBOHYD	319	AA;	AD2C3035ED45306 CRC64;
FT	SEQUENCE	319	AA;	AD2C3035ED45306 CRC64;
Alignment Scores:				
Pred. No.:	0.0165	Length:	319	
Score:	123.50	Matches:	78	
Percent Similarity:	34.22%	Conservative:	38	
Best Local Similarity:	23.01%	Mismatches:	112	
Query Match:	1.54%	Indels:	111	
DB:	1	Gaps:	15	
US-10-776-827-10 (1-4460) x TCB1_RABIT (1-319)				
Qy	954	ATGCTCTTACTGACAGCCAGGCATGCTCCACCTGT	---	992
Db	24	LeuThrMetThrProGluHisAlaArgProProThrThrThrAlaMetLeuTyrAspGly	43	
Qy	993	---CTCTGGCATACCCACATGCAAGACAAAGAACATTATCCATACATCTCAATATG	1049	
Db	44	PheLeuTrpProSerHisThrHisArgHisSerTyrLeuLeuHisProHisThrHisVal	63	
Qy	1050	GTTCCTCCAGTGTGGCATGCGACGTAAACACACACACACAAATTCAGGTACAGGTAC	1109	
Db	64	-----CysThrHisThrHisThrCys-ThrHisThrHisHis-	76	
Qy	1110	GTGGCAGATATATTCTGCTCATCAATGGTCAATGGCTATGATGACTTGTGACGGAAGT	1169	
Db	77	-----AlaSerThrHisValCysIleHisThrHisThrPheThrHisLeuCys	93	
Qy	1170	ACATTATCTACAGTCACAAAAATGTCTCATGGGAAAGCCCTTGGCAGATTTCAGACATAT	1229	
Db	93	eHisThrLeuThrHisAlaLeuThrLeuThrCysAlaProThrArgThrTyrAlaHis	112	
Qy	1230	ATACAAATTCGTACACGACGAGGCCCCATACACCATCTATTCCATAAACCACTCAGTT	1289	
Db	113	-----ThrArgAlaProThrHisValHisProHisLysProArgProAr	127	
Qy	1290	ACAGATGCATGCTTTCTCTATTCTTAACCTACACATAAACCTTTTCTGGAGTACTCATAT	1349	
Db	127	gGluLeuSerAlaAlaLeuLeuLeuProThrProLeuHisPheProGluAspLeuAlaAs	147	
Qy	1350	ATTGGACATTCACGAA	---	1378
Db	147	nValSerAlaPro-GlnValValValPheAspProSerGluAlaGluIleAsnLysThrG	167	
Qy	1379	-----CCCACCTTGTGTGCTTCATACAGAC	---	1405
Db	167	lnLysAlaThrLeuValCysLeuAlaLysaspPheTyrProAspHisValGluLeuSerT	187	
Qy	1406	-----ACACCAAGTTTCTGTGCTCTCGACCCCTCACCTTCTGCTGCGAA	1445	
Db	187	rpTpValAsnGlyLysGluValHisAsnGlyValSerThrAspProGlnProTyrLysG	207	
Qy	1446	GATGTTAAAGTGTGATGGTTCAAAATTCATTGAAGCTCTTTCTT	---	1492
Db	207	InAspProLysSerAspHisSerLysTyrCysLeuSerSerArgLeuArgValSerAlaA	227	
Qy	1493	--GTAACCTACGACAAAGTCGCTCTCAT	---	1544
Db	227	laPheTrpHisAsnPro---ArgAsnHisPheArgCysGlnValGlnPheGlyLeuT	246	
Qy	1545	TCCAAAGACTCTCTGTGAAACATTACCCCGCAACACTCAGCAAGTGCCTTCTCCA	1604	
Db	246	hrAspAspGluTrpThrTyrAsnSerSerLysProIleThr	---	260
Qy	1605	ACCAAGAACAAAGACTCTTGGTGGTGACTCTGAGAAATTTATGGAAGCCCACTCATTTA	1664	

```

Db 261 -----GlnAsnIleSerAlaHisThrArgG 269
QY 1665 TGTCAAGTCACTGCAACTGTGTACCTGTGCAATGTTTACAGATGGAAGGTCAGGAGAT 1724
Db 269 lyArgAlaaspCysGly----- 274
QY 1725 GCTACACCTGAGTAGGTATCTCTATATAACCAAGTTTCCAGCAGGGAAGGAACATGA 1784
Db 275 -----IleSerAlaSerTyrGlnGlnGly----- 283
QY 1785 CAATCATCAGTCGA-----GTCTCACAGAAGGCAACACTG 1819
Db 284 --ValIeuSerAlaThrValIeuTyrGluIleLeuLeuGlyLysAlaThrLeu 300
RESULT 3
V70K_TYMW
ID V70K_TYMW STANDARD; PRT; 628 AA.
AC P10357;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DE 01-AUG-1992 (Rel. 23, Last annotation update)
DE 69 kDa protein.
OS Turnip yellow mosaic virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tymoviridae;
OC Tymovirus.
OX NCBI_TaxID=12154;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88289359; PubMed=3399388;
RA Morch M.D., Boyer J.C., Haenni A.L.;
RT "Overlapping open reading frames revealed by complete nucleotide
RL sequencing of turnip yellow mosaic virus genomic RNA.";
RL Nucleic Acids Res. 16:6157-6173(1988).
CC -!- FUNCTION: Not known.
CC -!- SIMILARITY: TO 65 TO 70 kDa PROTEIN FROM OTHER TYMOVIRUSES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; X07441; CAA30321.1; ALT_SEQ.
DR PIR; S01955; S01955.
DR InterPro; IPR004935; Tymo.45kD.70kDa.
DR Pfam; PF03251; Tymo.45kD.70kD; 1.
SQ SEQUENCE 628 AA; 69195 MW; 9B01CB5ADPCEAC77 CRC64;
Alignment Scores:
Pred. No.: 0.0675 Length: 628
Score: 117.50 Matches: 101
Percent Similarity: 31.24% Conservative: 38
Best Local Similarity: 22.70% Mismatches: 141
Query Match: 1.46% Indels: 165
DB: 1 Gaps: 20
US-10-776-827-10 (1-4460) x V70K_TYMW (1-628)
QY 2748 ATCCCTCTGTGCATAGAGAGTTTGTCTCATGGGCGCT-----CTGGCTGTGCCCTTC 2798
Db 40 LeuProMetValHisSerGluGlyThrSerThrProThrGlnLeuLeuArgHisProAsn 59
QY 2799 ACATTAACAGATAACTTCCCATCTGCTGCACCAACCCAGGAGTGTGGAAGACATCTCC 2858
Db 60 IleTrpPheGlyAsnLeuProProProProArgArgProGlnAspAsnArgAspPheSer 79
QY 2859 CCAGAACTGCCACTGCTC----- 2876
Db 80 ProLeuHisProLeuAlaPheProGlyHisHisSerGlnLeuLeuArgHisValHisGluThr 99
QY 2877 -----ACCAGACAAAGCTGCCCTTCCTGTCTCCACCTCTCAGTCCCCCTAGATGGATG 2930

```

```

Db 100 GlnGlnValGlnGlnThrCysProGlyLysLeuLysLeuSer----- 113
QY 2931 GCTGGGGAGAGGTGGAGGCTGACAGCTGACAGCTAGTGTGAGATATGATCTAGGAGGGCG 2990
Db 114 -----GlyAla 115
QY 2991 GATCACCGGATTCGGGACCATACAAGTAACATGCT-----TTCCATCGCAACTGCTTGT 3044
Db 116 GluGluLeuProProAlaProGlnArgGlnHisSerLeuProLeuHis-IleThrArgPr 135
QY 3045 CTCG-----TTTGAATTAGACAGCAGTCAGTT----- 3072
Db 135 oSerArgPheProHisHisPheHisAlaArgArgProAspValLeuProSerValProAs 155
QY 3073 -----GTCATTGCCATGACCAAGGCTCTATCT----- 3099
Db 155 phiGlyProValLeuThrGluThrIleProArgThrSerValArgGlnProArgSerAl 175
QY 3100 -----CCAGGCAAAATGCTCCTGCTGCTCTCTAAT----- 3129
Db 175 aThrArgGlyProSerPheArgProIleLeuLeuProLysValHisValHisAspAs 195
QY 3130 -----CCAATGGACTTGTCTCTACCCAGGATGAACACCCAGAACTCACT----- 3177
Db 195 pProHisSerSerLeuArgProArgGlySerArgSerArgGlnLeuGlnProThrVa 215
QY 3178 -----TCTCAGTCACCTTCCACAGCCGATGAC 3203
Db 215 largArgProLeuLeuAlaProAsnGlnPheHisSerProArgGlnProProLeuSe 235
QY 3204 TCAGAAGAGCCAAACCCAGATGGGGCGCTCTCTTTTC----- 3240
Db 235 raspAspProGlyIleLeuGlyProArgProLeuAlaProHisSerThrArgAspProPr 255
QY 3241 -----CCCATCAGACTCCCTGCAAC-----CTTCTCTGGCG 3275
Db 255 oProArgProIleThrProGlyProSerAsnThrHisAspLeuArgProLeuSerVal 275
QY 3276 TAACTAGAGAGTCCAGTCAGGATAGCCCTAAACCTTTTGTAAATAAACAGCTGCA 3335
Db 275 uProArgThrSerProArgArgGly-----LeuLeuProAsnProArgArgHi 291
QY 3336 TGAAGAGGCTTAAGGCCATTTGTGATATCCACTCTCTTCTTCCACTCTCTCTCTCATCT 3395
Db 291 sArgThrSerThr-GlyHis----- 297
QY 3396 TTTTCTCCATGTTTATGCTTCTCTGATTCCTCTTCTGCTGCTGACACAGCAGCC----- 3451
Db 298 -----IleProProThrThrThrSerArgProThrGlyP 309
QY 3452 --CCAGCCCTTTATTCCTCTCCATTTTCACTCTTCCAGCCTCTGCTCCCTGAACCTGCCAC 3509
Db 309 roProSerArgLeuGlnArgPro-ValHisLeuTyrGlnSer-SerProHisThrProAs 328
QY 3510 TGGCAACCCATGGGACCTCAGCAGACAGACTGTGACTCATCTGGGG----- 3558
Db 328 nPheArgProSerSerIleArg---LysAspAlaLeuLeuGlnThrGlyProArgLeuGl 347
QY 3559 -----AGGTAAGTTTCCGGGGGACAAAAA 3584
Db 347 yHisLeuGluArgLeuGlyGlnProAlaAsnLeuArgThrSerGluArgSerProPro-- 366
QY 3585 ATGATTCCTAAAGAAAGAGGCTTCTAGACACAGCAGCTCCAGAAAGACATCCCTTAGG 3644
Db 367 -----ThrIysArgArgLeuProArgSerSerGlu---ProAsnArg--LeuProLys 382
QY 3645 CCTGGACTTCTGACAGCTTTTGAAGAGCTCCGGACCGGACGCCAGAGAGAGGCTTTCCCC 3704
Db 383 Pro-----LeuProGluAlaThrLeuAlaProSerTyrArgHisArgArgProTyrPro 400
QY 3705 ATTGCTCTTCTTCCCATGTCTCAATGGATTCATGTTTCTTTTCTTGGGGGGAGCAGGG 3764

```

```

Db 401 LeuLeuProAsnProProAlaAala---LeuProSerIleAlaTyrThrSerSerArgGly 419
QY 3765 AGG 3767
Db 420 Lys 420
RESULT 4
W70K_TYMCV
ID W70K_TYMCV STANDARD; PRT; 628 AA.
AC P28478;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DE 69 kDa protein.
OS Turnip yellow mosaic virus (isolate TYMC).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tymoviridae;
OC Tymovirus.
OX NCBI_TaxID=31751;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92119261; PubMed=1731998;
RA Dreher T.W., Branson K.L.;
RT "Genomic RNA sequence of turnip yellow mosaic virus isolate TYMC, a
RL cDNA-based clone with verified infectivity.";
RL Plant Mol. Biol. 18:403-406(1992).
CC -!- FUNCTION: Not known.
CC -!- SIMILARITY: TO 65 TO 70 kDa PROTEIN FROM OTHER TYMOVIRUSES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL; X16378; CAA34414.1; -.
DR PIR; S19150; S19150.
DR InterPro; IPR004935; Tymo_45_70kDa.
DR Pfam; PF03251; Tymo_45kDa_70kDa; 1.
SQ SEQUENCE 628 AA; 69057 MW; 0DF0C64B8ECBC6DC CRC64;

Alignment Scores:
Pred. No.: 0.168 Length: 628
Score: 113.00 Matches: 103
Percent Similarity: 31.53% Conservative: 37
Best Local Similarity: 23.20% Mismatches: 139
Query Match: 1.41% Indels: 165
DB: 1 Gaps: 22

US-10-776-827-10 (1-4460) x W70K_TYMCV (1-628)
QY 2748 ATCCCTTTGTGCATAGAGATTGTTCATGGGCTC-----CTGGCTGTGCCCTTC 2798
Db 40 LeuProMetValHisSerGluGlyThrSerThrProThrGlnLeuLeuArgHisProAsn 59
QY 2799 ACATAACAGATAACTTGCCTATCTGCTGCACCAACCCAGGGATGTGGAGACATCTCC 2858
Db 60 IleTrpPheGlyAsnLeuProProProArgArgProGlnAspAsnArgAspPheSer 79
QY 2859 CCACAACTGCCACTGCTC----- 2876
Db 80 ProLeuHisProLeuValPheProGlyHisHisSerGlnLeuArgHisValHisGluThr 99
QY 2877 -----ACCAGGACAGAGTGCCTTCTGCTCTCCACCTCTCAGTCCCTAGATGGATG 2930
Db 100 GlnGlnValGlnGlnThrCysProGlyLysLeuLysLeuSer----- 113
QY 2931 GCTGGGGAGAGGTGAGGCTGACAGCTGAGAGCTAGTGTCTAGATATGATCTAGAGGGCG 2990
Db 114 -----GlyAla 115
QY 2991 GATCACCAGGATCCGGGACCACATACAGTAACATGCT-----TTCCATGGCACTGCTG 3044

```

```

Db 116 GluGluLeuProProAlaProGlnArgGlnHisSerLeuProLeuHis-IleThrArgPr 135
QY 3045 CTGG-----TTTGAATTAAAGACAGCAGTCAGTT----- 3072
Db 135 oSerArgPheProHisPheHisAlaArgArgProAspValLeuProSerValProAs 155
QY 3073 -----GTCATTGCCATGACAAAGCCCTCTATCT----- 3099
Db 155 pHisGlyProValLeuAlaGluThrLysProArgThrSerValArgGlnProArgSerAl 175
QY 3100 -----CCAGGCACAATGTCCTGCTGCTCTCTAAT----- 3129
Db 175 aThrArgGlyProSerPheArgProIleLeuLeuProLysValValHisValHisAspAs 195
QY 3130 -----CCAATGGACTGCTCTCACCAGGGATGAACACCCAGAACTCACT----- 3177
Db 195 pProProHisSerSerLeuArgProArgLysArgSerArgSerGlnLeuGlnProThrVa 215
QY 3178 -----TCTCAGTCACCTTCCACAGCCGATGAC 3203
Db 215 lArgArgProLeuLeuAlaProAsnGlnPheHisSerProArgGlnProProLeuSe 235
QY 3204 TCAGAGAGCCAAACCCAGATGGGCTCTCTTTC----- 3240
Db 235 rAspAspProGlyIleLeuLeuGlyProArgProLeuAlaProAsnSerThrArgAspProPr 255
QY 3241 -----CCCATCAGAGCTCCCTGACAAAC-----CTTTCTCTGGCG 3275
Db 255 oProArgProIleThrProGlyProSerAsnThrHisGlyLeuArgProLeuSerValLe 275
QY 3276 TAACTAGAGGAGTCCAGGTGACAGATAGGCCCTTAAACGTTTGTAAATAACAGGTGCA 3335
Db 275 uProArgAlaSerProArgArgGly-----LeuLeuProAsnProArgArgHi 291
QY 3336 TGAAGAGCCCTAAGGCCATTGTTGATATCCACTCTCTCTTCTTCCACTT-----CCTTC 3389
Db 291 sArgThrSerThrGlyHisIleProProThrThrThrSerArgProThrGlyProProSe 311
QY 3390 TCATCTTTTCTCCATGTTTATGCTTCTGATTCCTCTCTGCTGCTGCACAGACCAG 3449
Db 311 rArgLeuGlnArgProVal-----HisLeuTyrGlnSerSe 323
QY 3450 CCCAGCCCTTATTCTCTCCATTTTCACTCTCCAGCTC----- 3492
Db 323 rPro-----HisThrProAsnPheArgProSerSerIleArgLysAspAlaLeuLe 340
QY 3493 -----TGTCCTGNACTGCCACTGGCAACCCATG 3521
Db 340 uGlnThrGlyProArgLeuGlyHisLeuGluCys-----LeuGlyGlnProAl 356
QY 3522 GGACCTCAGGACCCAGAGACTGCTTGACTCATCTGGGAGGGTAAAGTTACGGGGGACAAA 3581
Db 356 aAsnLeuArgThr-----SerGluArgSerProPr 366
QY 3582 AAAATGATTCCTAAAGAAAGAGGCTTCTAGACAGACAGCTCCAGAAAGACATCCCT 3641
Db 366 o-----ThrLysArgArgLeuProArgSerSerGlu---ProAsnArg--LeuPro 381
QY 3642 AGGCTCGACTTCTGACAGCTTTAGCCAGCTCCGACCGCAGCCAGCAGAGAGGCTTTC 3701
Db 382 LysPro-----LeuProGluAlaThrLeuAlaProSerTyrArgHisArgSerTyr 399
QY 3702 CCCATTGCTCTTCTCCCATTTGCTCAATGGATTCCATGTTTCTTTTCTGGGGGGAGCA 3761
Db 400 ProPheLeuProAsnProProAlaAlaAala---LeuProSerIleAlaTyrThrSerArg 418
QY 3762 GGGAGG 3767
Db 419 GlyLys 420
RESULT 5
HNFA_HUMAN

```

ID HNFA HUMAN STANDARD; PRT; 631 AA.
 AC P20823; Q99861;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Hepatocyte nuclear factor 1-alpha (HNF-1A) (Liver-specific
 DE transcription factor LF-B1) (LFB1) (Transcription factor-1) (TCF-1).
 GN TCF1 OR HNF1A.
 OS Homo sapiens (Human).
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=91184801; PubMed=1707031;
 RA Buch I., Galcheva-Gargova Z., Mattei M.-G., Simon-Chazottes D.,
 RA Guenet J.-L., Cereghini S., Yaniv M.;
 RT "Cloning of human hepatic nuclear factor 1 (HNF1) and chromosomal
 RT localization of its gene in man and mouse.";
 RL Genomics 8:155-164(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=94038905; PubMed=7900999;
 RA Bach I., Yaniv M.;
 RT "More potent transcriptional activators or a transdominant inhibitor
 RT of the HNF1 homeoprotein family are generated by alternative RNA
 RT processing.";
 RL EMBO J. 12:4229-4242(1993).
 RN [3]
 RP SEQUENCE FROM N.A., AND VARIANT MODY3 LEU-447.
 RX MEDLINE=97100943; PubMed=8945470;
 RA Yamagata K., Oda N., Kaisaki P.J., Menzel S., Furuta H.,
 RA Vaxillaire M., Southern L., Cox R.D., Lathrop G.M., Boriraj V.V.,
 RA Chen X., Cox N.J., Oda Y., Yano H., le Beau M.M., Yamada S.,
 RA Nishigori H., Takeda J., Fajans S.S., Hattersley A.T., Iwasaki N.,
 RA Hansen T., Pedersen O., Polonsky K.S., Turner R.C., Velho G.,
 RA Chevre J.-C., Froguel P., Bell G.I.;
 RT "Mutations in the hepatocyte nuclear factor-1alpha gene in maturity-
 RT onset diabetes of the young (MODY3).";
 RL Nature 384:455-458(1996).
 RN [4]
 RP VARIANTS MODY3 ARG-107; TRP-131; MET-260 AND HIS-272.
 RX MEDLINE=97309269; PubMed=9166684;
 RA Glucksman M.A., Lehto M., Tayber O., Scotti S., Berkemeter L.,
 RA Pulido J.C., Wu Y., Nir W.-J., Fang L., Markel P., Munnelly K.D.,
 RA Goranson J., Orho M., Young B.M., Whitacre J.L., McMenimen C.,
 RA Wantman M., Tuomi T., Warram J., Forablon C.M., Carlsson M.,
 RA Rosenzweig J., Kennedy G., Duyk G.M., Krolewski A.S., Groop L.C.,
 RA Thomas J.D.;
 RT "Novel mutations and a mutational hotspot in the MODY3 gene.";
 RL Diabetes 46:1081-1086(1997).
 RN [5]
 RP VARIANTS MODY3 HIS-12; GLN-131; GLN-205 AND CYS-263, AND VARIANT NIDDM
 ASP-191.
 RX MEDLINE=97431548; PubMed=9287053;
 RA Iwasaki N., Oda N., Ogata M., Hara M., Hinokio Y., Oda Y.,
 RA Yamagata K., Kanematsu S., Ohgawara H., Omori Y., Bell G.I.;
 RT "Mutations in the hepatocyte nuclear factor-1alpha/MODY3 gene in
 RT Japanese subjects with early- and late-onset NIDDM.";
 RL Diabetes 46:1504-1508(1997).
 RN [6]
 RP VARIANT NIDDM MET-254, AND VARIANTS LEU-27 AND ASN-487.
 RX MEDLINE=97431550; PubMed=9287055;
 RA Yamada S., Nishigori H., Onda H., Takahashi K., Kitano N.,
 RA Morikawa A., Takeuchi T., Takeda J.;
 RT "Mutations in the hepatocyte nuclear factor-1alpha gene (MODY3) are
 RT not a major cause of late-onset NIDDM in Japanese subjects.";
 RL Diabetes 46:1512-1513(1997).
 RN [7]
 RP VARIANTS IDDM HIS-272 AND GLY-583.
 RX MEDLINE=97458988; PubMed=9313763;
 RA

Yamada S., Nishigori H., Onda H., Utsugi T., Yanagawa T., Maruyama T.,
 Onigata K., Nagashima K., Nagai R., Morikawa A., Takeuchi T.,
 Takeda J.;
 RT "Identification of mutations in the hepatocyte nuclear factor (HNF)-1-
 alpha gene in Japanese subjects with IDDM.";
 RL Diabetes 46:1643-1647(1997).
 RN [8]
 RP VARIANTS MODY3, AND VARIANT ATYPICAL DIABETES SER-574.
 RX MEDLINE=98052398; PubMed=9392505;
 RA Boutin P., Chevre J.-C., Hani E.H., Gomis R., Pardini V.C.,
 RA Guillausseau P.-J., Vaxillaire M., Velho G., Froguel P.;
 RT "An automated fluorescent single-strand conformation polymorphism
 RT technique for screening mutations in the hepatocyte nuclear
 RT factor-1alpha gene (maturity-onset diabetes of the young).";
 RL Diabetes 46:2108-2109(1997).
 RN [9]
 RP VARIANTS MODY3 GLN-131; GLN-229; GLY-241 AND HIS-272.
 RX MEDLINE=97184312; PubMed=9032114;
 RA Kaisaki P.J., Menzel S., Lindner T., Oda N., Rjasanowski I., Sahn J.,
 RA Meincke G., Schulze J., Schmechel H., Petzold C., Ledermann H.M.,
 RA Sachse G., Boriraj V.V., Menzel R., Kerner W., Turner R.C.,
 RA Yamagata K., Bell G.I.;
 RT "Mutations in the hepatocyte nuclear factor-1alpha gene in MODY and
 RT early-onset NIDDM: evidence for a mutational hotspot in exon 4.";
 RL Diabetes 46:528-535(1997).
 RN [10]
 RP VARIANTS MODY3 THR-129; TRP-131; TRP-159; LEU-519 AND ILE-620.
 RX MEDLINE=97230240; PubMed=9075818;
 RA Frayling T.M., Bulman M.P., Ellard S., Appleton M., Dronsfield M.J.,
 RA Mackie A.D., Baird J.D., Kaisaki P.J., Yamagata K., Bell G.I.,
 RA Bain S.C., Hattersley A.T.;
 RT "Mutations in the hepatocyte nuclear factor-1alpha gene are a common
 RT cause of maturity-onset diabetes of the young in the U.K.";
 RL Diabetes 46:720-725(1997).
 RN [11]
 RP VARIANTS MODY3 ASN-128; TYR-143 AND LEU-447.
 RX MEDLINE=97230241; PubMed=9075819;
 RA Hansen T., Elberg H., Rouard M., Vaxillaire M., Moeller A.M.,
 RA Rasmussen S.K., Fridberg M., Urhammer S.A., Holst J.J., Almind K.,
 RA Schwald S.M., Hansen L., Bell G.I., Pedersen O.;
 RT "Novel MODY3 mutations in the hepatocyte nuclear factor-1alpha gene:
 RT evidence for a hyperexcitability of pancreatic beta-cells to
 RT intravenous secretagogues in a glucose-tolerant carrier of a P447L
 RT mutation.";
 RL Diabetes 46:726-730(1997).
 RN [12]
 RP VARIANTS LEU-27; VAL-98 AND ASN-487.
 RX MEDLINE=97278987; PubMed=9133564;
 RA Urhammer S.A., Fridberg M., Hansen T., Rasmussen S.K., Moeller A.M.,
 RA Clausen J.O., Pedersen O.;
 RT "A prevalent amino acid polymorphism at codon 98 in the hepatocyte
 RT nuclear factor-1alpha gene is associated with reduced serum C-peptide
 RT and insulin responses to an oral glucose challenge.";
 RL Diabetes 46:912-916(1997).
 RN [13]
 RP VARIANT NIDDM GLN-583, AND VARIANTS LEU-27; VAL-98 AND ASN-487.
 RX MEDLINE=97266232; PubMed=9112026;
 RA Urhammer S.A., Rasmussen S.K., Kaisaki P.J., Oda N., Yamagata K.,
 RA Moeller A.M., Fridberg M., Hansen L., Hansen T., Bell G.I.,
 RA Pedersen O.;
 RT "Genetic variation in the hepatocyte nuclear factor-1 alpha gene in
 RT Danish Caucasians with late-onset NIDDM.";
 RL Diabetologia 40:473-475(1997).
 RN [14]
 RP VARIANTS MODY3 CYS-122; PHE-142 AND GLN-159.
 RX MEDLINE=97253393; PubMed=9097962;
 RA Vaxillaire M., Rouard M., Yamagata K., Oda N., Kaisaki P.J.,
 RA Boriraj V.V., Chevre J.-C., Boccio V., Cox R.D., Lathrop G.M.,
 RA Dussoix P., Philippe J., Timsit J., Charpentier G., Velho G.,
 RA Bell G.I., Froguel P.;
 RT "Identification of nine novel mutations in the hepatocyte nuclear
 RT factor 1 alpha gene associated with maturity-onset diabetes of the
 RT young (MODY3).";


```

Db      537 ThrProThrLysGlnValPheThrSerAspThr-----GluAlaSerSerGluSerGly 554
QY      1254 CCCATACACCATCTATTCCATAAACCACCTCAGGTTACAGATGCATGCTTTCCATTCTT 1313
Db      555 LeuHisThrProAlaSerGlnAlaThrThrLeuHisValProSerGlnAspProAlaGly 574
QY      1314 -----AACTCTACACATAACTTTTACTGGAAGTACTCATATTGACATTCC 1361
Db      575 IleGlnHisLeuGlnProAlaHisArgLeuSerAlaSerProThrValSerSerSer 594
QY      1362 AGCAACCTCTACAGTCC 1379
Db      595 LeuValLeuTyrGlnSer 600

RESULT 6
V70K_OYMW
ID V70K_OYMW STANDARD; PRT; 597 AA.
AC P20130;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-FEB-1991 (Rel. 17, Last annotation update)
DE 66 kDa protein.
OS Ononis yellow mosaic virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tymoviridae;
OC Tymovirus.
OX NCBI_TaxID=12153;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90021186; PubMed=2800337;
RA Ding S.W., Keese P., Gibbs A.;
RT "Nucleotide sequence of the ononis yellow mosaic tymovirus genome.";
RL Virology 172:555-563 (1989).
CC 1- FUNCTION: Not known.
CC 2- SIMILARITY: TO 65 TO 70 kDa PROTEIN FROM OTHER TYMOVIRUSES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; J04375; AA446795.1; -
DR PIR; JQ0107; JQ0107.
DR InterPro; IPR004935; Tymo_45_70kDa.
DR Pfam; PF03251; Tymo_45kd_70kd; 1.
SQ SEQUENCE 597 AA; 65985 MW; 8A0BE4C2A7D329F8 CRC64;

Alignment Scores:
Pred. No.: 0.203 Length: 597
Score: 112.00 Matches: 96
Percent Similarity: 33.25% Conservative: 45
Best Local Similarity: 22.64% Mismatches: 121
Query Match: 1.40% Indels: 163
DB: 1 Gaps: 24

US-10-776-827-10 (1-4460) x V70K_OYMW (1-597)
QY 2656 CCTCTCACCTATATGCGCACTAAAGCTGCTTAAGCACTCAGGCTC----- 2700
Db 187 ProLysProLeuHisLeuHisAsnPheArgGlnHisSerLeuHisSerLeuHisSerArgSer 206
QY 2701 -----CCACTCATCAACCCCTTTGACACAGAGAAAGACAC 2736
Db 207 ProCysArgLysLeuArgProThrProArgCysAsnGlnLeuAlaGlnAlaGlnHis 226
QY 2737 -----TCTGTTCTCTATCCCTTTGTACATA 2763
Db 227 ProLeuProSerSerLysProLeuSerLeuGlnAlaGly11LeuGlyProCysProLeu 246
QY 2764 GAGAGTTTGTCTATGGGCTCTGGCTGTGCTTTCACATACAGATAACTTGGCATCTG 2823

```

```

Db      247 -----ProProHisAsnLys----- 251
QY      2824 CTGTCAACCAAAACCCAGGATGTGAAGACATCT-----CCCAACAACCTGCCAC 2871
Db      252 ArgSerProSerProAlaValIleArgLysThrAlaGlyIleLeuProHisProLysLeu 271
QY      2872 TGCTCACCAAGCAAGCTGCCCTTCTCTCCACTCTCAGTCCCTCTAGATGGATGG 2931
Db      272 ProProSerSerArgGlyHisLeuProSerSerThrSerSerSer----- 286
QY      2932 CTGGGGAGAGTGGAGGCTGACAGCTGACAGTGTGTGATATCATATCATAGAGGGCGG 2991
Db      287 -----SerProArgSerAsnArgGlyValArg 295
QY      2992 ATCACCGGATCGGGACCATACAACTAATGTTTCCATG--GCAACTGCTTCG-- 3045
Db      296 CysSer---ValHisLeuHisLysSerArgSerAsnSerGlnAspLeuArgSerCysArg 314
QY      3046 -----TCGTTTGAATTAAAGACAGCAGTCAGTTGTTCATTGCCATGACAAAGGCTCTA--- 3096
Db      315 ValArgSerAsnSerLeuGlnGlnThrProIleLeuMetGlyHisPheLysSerLeuGly 334
QY      3097 ---TCTCCAGGCACAATGTCCCTGCTGTCTCTAATCCA-----ATG 3135
Db      335 GlnSerProAsnLeuArgSerPheGluArgProArgProThrArgArgSerLeuArgLeu 354
QY      3136 GACTTGTCTCTACCCCGGATGAAACACCCAGAAACTCACTTCTCAG----- 3183
Db      355 LeuProLeuSerProGlnLysValProThrValHisValProThrHisGlnSerGly 374
QY      3184 -----TCATTCCACCGCATGATCAGAAAGAGCCAAACCCAGATGGGGCC 3231
Db      375 HisLysGlyProSerLeuProArgProHisSerProSerArgGlnThrHisHisala--- 393
QY      3232 TCTCTTTTCCCATCACAGACTCCCTCCACAACCTTCTCGCGTAACTAGAGGAGTCCC 3291
Db      394 -----ArgLeuPro 396
QY      3292 AGTCAGGATAGGCCCTAAACGTTTGTAAATAAACAGGTG-----CATGAAAGGAGC 3345
Db      397 His-----SerLysArgValSerLeuProAsnSerValLeuHisHisaspArgPro 413
QY      3346 CTAGGCAATGTGTGATATCCACTCTCTTCTTTCACATT-----CCTTCTCATCTT 3396
Db      414 LysArgProIleHisPheGlySer-----PheProIleAsnValAlaProSerHisLeu 431
QY      3397 TTTCTCCATGTTTATGCTTCTCTGATTCCTCTGATTCCTCTGCTGCACACAGCCCGCCAGC 3456
Db      432 Leu-----ProArg 434
QY      3457 CTTTATTCTCTCTCCTTCTCCTTCTCCTTCTCCTTCTCCTTCTCCTTCTCCTTCTCCTTCT 3513
Db      435 LysLeuTrpSerArgAlaSerSerProProThrCysSerProThrThrSerAsnHisGly 454
QY      3514 AACCCATGGGAC---CTCAGGACCCAGAGACTGCTTTGACTCATCTCTGGGGAGGTAACTTCA 3570
Db      455 HisProGluGluAlaLeuArg----- 461
QY      3571 CGGGGGACAAAAAATGTTCTTAAGAGAGGCTTCTTAGACACGACGACGCG----- 3623
Db      462 -----PheLeuProLysAsnLeuProGlnHisCysGlnMetAlaLeu 475
QY      3624 -----TCCAGAAACACATCCCTAGGCTGGAGCTTCTGAGCAGCTTTAGC 3668
Db      476 MetGluAsnTyrCysSerHisPheSerSerPro-----SerSerSerVal 490
QY      3669 CAGGCTCCGAGGACGACGACGAGGAGGCTTTTCCCATTTGCTCTCTTCCCATTTGCTCAA 3728
Db      491 SerPheProGluAspHisGlnSerSerLeuProIleSer-----ThrArg 506
QY      3729 TGGATTCCATGT 3740

```


QY 1230 ATACAATTTCTTACACAGCAGGCCCCATACACCATCTTATCCATAAACCACTCAGGTT 1289
 Db 110 -----ProSerProSerThrProProThrProSerProPro----- 121
 QY 1290 ACAGATGATGCTTTCTTATTTCTTAACTTACACATAAACTTTTACTGGAAGTACTCAT 1349
 Db 122 -----CysCysHisPro---ArgLeuSerLeuHisArgProAlaLeuGluAspLeuLeu 138
 QY 1350 ATTGGACATTTCCAGCAACCTG----- 1370
 Db 139 LeuGlySerGluAlaAsnLeuThrCysThrLeuThrGlyLeuArgAspAlaSerGlyVal 158
 QY 1371 -----CTACAGTCCCCACCTTGTGTGTC 1394
 Db 159 ThrPheThrTrpThrProSerSerGlyLeuSerAlaValGluGlyProProGluArgAsp 178
 QY 1395 TTGATACAGACACCAACCTTTCTGTCTCTGACCCCTTCACTGTGCGCAAGATGTTAA 1454
 Db 179 LeuGlySerGlyCysTySerValSerSerValLeuPro---GlyCysAlaGluProTrpAs 197
 QY 1455 AGTGTGATGTTCAAAATTCATGAAGCTCTTTCTTGTAACTCATGACAAAGTCCGTC 1514
 Db 197 n-----HisGlyLysThrPheThrCysThrAlaAlaLysProGluSer 211
 QY 1515 CTCATGTCCTGACGAGGCTTTAATGTGATCA-----AGACCTCTCTGTGCA 1562
 Db 211 rLysThrProLeuThrAlaThrLeuSerLysSerGlyAsnMetPheArgProGluValHi 231
 QY 1563 AACATTACCCCGCAACCACTCAG 1587
 Db 231 sLeuLeuProProProSerGluGlu 239
 RESULT 9
 Z409 HUMAN STANDARD; PRT; 862 AA.
 AC Q9UP6;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Zinc finger protein 409.
 GN ZNF409 OR KIAA1056.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=93397452; PubMed=10470851;
 RA Kikuno R., Nagase T., Ishikawa K.-I., Hirose M., Miyajima N.,
 RA Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XIV.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro.";
 RL DNA Res. 6:197-205 (1999).
 CC -!- FUNCTION: May function as a transcription factor.
 CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -!- SIMILARITY: Contains 6 C2H2-type zinc fingers.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. Usage by and for commercial
 CC or send an email to license@isb-sib.ch.
 CC
 CC EMBL; AB028979; BAA8308.1; --
 DR HSSP; P15822; 3ZNF.
 DR InterPro; IPR007087; Znf_C2H2.
 DR Pfam; PF00096; zF-C2H2; 5.
 DR SMART; SM00355; Znf_C2H2; 6.

DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 2.
 KW PROSITE; PS0157; ZINC_FINGER_C2H2_2; FALSE NEG.
 KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
 KW Nuclear protein; Repeat.
 FT ZN_FING 237 259 C2H2-TYPE 1.
 FT ZN_FING 453 476 C2H2-TYPE 2.
 FT ZN_FING 508 532 C2H2-TYPE 3.
 FT ZN_FING 569 593 C2H2-TYPE 4.
 FT ZN_FING 758 782 C2H2-TYPE 5.
 FT ZN_FING 821 845 C2H2-TYPE 6.
 FT DOMAIN 605 710 PRO-RICH.
 SQ SEQUENCE 862 AA; EB6BFF9A0B939217 CRC64;
 Alignment Scores:
 Pred. No.: 0.506 Length: 862
 Score: 108.00 Matches: 97
 Percent Similarity: 28.25% Conservative: 53
 Best Local Similarity: 18.27% Mismatches: 158
 Query Match: 1.35% Indels: 223
 DB: Gaps: 1
 US-10-776-827-10 (1-4460) x Z409_HUMAN (1-862)
 QY 26 GGGCTTGAGAGAGCCCTTCCGCGACCGAAGATTGTT----- 64
 Db 223 GlyProMetGlyAsnSerGlyGlyAsnHisValAlaValPheTrpLeuCysLeuLeuCys 242
 QY 65 -----CCCATTTTGGAGATGAAGAACTCGAGACTCAAAGCAGCTGA 106
 Db 243 ArgLeuGlyPheSerLysProGlnAlaPheMetAspHisThrGlnSerHisGlyValLys 262
 QY 107 GTGACCTTCCCAAGGACACACACTGAACCTGGCGGTGATCAGATCTGAATGCACAGGCG 166
 Db 263 LeuThr---ProAlaGlnTyrGlnGlyLeuSerGlySerProAlaValLeuGlnGluGly 281
 QY 167 -----GGGTGTTTCAGGATGTTTACTACGTTGAACGTGACCTCCAGGAAGCAGTTCT 220
 Db 282 AspGluGlyCysLysAlaLeuLeuSerPheLeuGluProLysLeuProAlaArgProSer 301
 QY 221 GCGCGAGATCCCTGACACAGCAAGCAAGAACTAACCTGGAAGGA----- 266
 Db 302 -SerAspIleProLeuAspAsnSerSerThrValAsnMetGluAlaAsnValAlaGlnTh 321
 QY 267 -----GGCTCCCA-----AGCTGGTGGCCATTGCTGCTGTGTGAGGTGCT 313
 Db 321 rGluAspGlyProProGluAlaGluValGlnAlaLeuLeuLeuLeuAspGluValMe 341
 QY 314 GTCAGTGGCATGCCAAACCCA-----AAGCTGGA 343
 Db 341 tAlaLeuSerProProSerProProThrAlaThrTrpAspProSerProThrGlnAlaLys 361
 QY 344 AGAGGAATRAATTAAGTGTCAAGTGTCAAGTGTGATCTCTTTCAGTCAAGACCT----- 395
 Db 361 sGluSerProValAlaAlaGlyGlu-----AlaGlyProAspTrpPh 375
 QY 396 -----GCTTGAAGCCGAGAGGGTCTCTGCGCCCTAA 427
 Db 375 eProGluGlyGlnGluAspGlyGlyLeuCysProProLeuAsnGlnSerSerProTh 395
 QY 428 TCTAGCAACACCATGAGAGATCAGTGCCTTCTCAGCTCTATCTGGGACACCATCT 487
 Db 395 r---SerLysGluGlyGly-----ThrLeuProAlaProValGlySerProGl 410
 QY 488 TGACCAACACCAAGAGGATCTTACACACCATCTCTGCTGGGAGTCTCTCTGGGCTGC 547
 Db 410 uAsp-----ProSerAspProProGlnProTy 419
 QY 548 CACTCTTGTGATCATCATCTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 607
 Db 419 rArgLeuAlaAspAspTyrThrPro----- 427
 QY 608 GCAAGAGGGGCCAGCAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 667

RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilms L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.;
RA "The DNA sequence and comparative analysis of human chromosome 20.";
RL Nature 414:865-871(2001).
RN [7]
RP INTERACTION WITH NCOA6IP.
RX MEDLINE=21417756; PubMed=11517327;
RA Zhu Y.-J., Qi C., Cao W.-Q., Yeldandi A.V., Rao M.S., Reddy J.K.;
RT "Cloning and characterization of PIMT, a protein with a
RT methyltransferase domain, which interacts with and enhances nuclear
RT receptor coactivator PRIP function.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:10380-10385(2001).
RN [8]
RP INTERACTION WITH RBM14.
RX MEDLINE=21423995; PubMed=11443112;
RA Iwasaki T., Chin W.W., Ko L.;
RT "Identification and characterization of RRM-containing coactivator
RT activator (CoAA) as TRBP-interacting protein, and its splice variant
RT as a coactivator modulator (CoAM)." ;
RL J. Biol. Chem. 276:33375-33383(2001).
RN [9]
RP INTERACTION WITH HRMTLL1.
RX MEDLINE=22151129; PubMed=12039952;
RA Qi C., Chang J., Zhu Y., Yeldandi A.V., Rao S.M., Zhu Y.-J.;
RT "Identification of protein arginine methyltransferase 2 as a
RT coactivator for estrogen receptor alpha." ;
RL J. Biol. Chem. 277:28624-28630(2002).
RN [10]
RP INTERACTION WITH MLL3 AND THE ASCOM COMPLEX.
RX TISSUE=Cervical carcinoma;
RX MEDLINE=22371496; PubMed=12482968;
RA Goo Y.-H., Sohn Y.-C., Kim D.-H., Kim S.-W., Kang M.-J., Jung D.-J.,
RA Kwak E., Barlev N.A., Berger S.L., Chow V.T., Roeder R.G.,
RA Azorsa D.O., Meltzer P.S., Sun P.-G., Song E.J., Lee K.-J., Lee Y.C.,
RA Lee J.W.;
RT "Activating signal cointegrator 2 belongs to a novel steady-state
RT complex that contains a subset of trithorax group proteins." ;
RL Mol. Cell. Biol. 23:140-149(2003).
RN [11]
RP MUTAGENESIS OF 883-THR-GLU-894, AND PHOSPHORYLATION.
RX MEDLINE=21635582; PubMed=11773444;
RA Ko L., Cardona G.R., Iwasaki T., Bramlett K.S., Burris T.P.,
RA Chin W.W.;
RT "Ser-884 adjacent to the LXXLL motif of coactivator TRBP defines
RT selectivity for ERs and TRs" ;
RL Mol. Endocrinol. 16:128-140(2002).
CC -1- FUNCTION: Nuclear receptor coactivator that directly binds nuclear
CC receptors and stimulates the transcriptional activities in a
CC hormone-dependent fashion. Coactivates expression in an agonist-
CC and AP2-dependent manner. Involved in the coactivation of
CC different nuclear receptors, such as for steroids (GR and ERs),
CC retinoids (RARs and RXRs), thyroid hormone (TRs), vitamin D3 (VDR)
CC and prostanoids (PPARs). Probably functions as a general
CC coactivator, rather than just a nuclear receptor coactivator. May
CC also be involved in the coactivation of the NF-kappa-B pathway.
CC May coactivate expression via a remodeling of chromatin and its
CC interaction with histone acetyltransferase proteins.
CC -1- SUBUNIT: Monomer and homodimer. Interacts with RNPC2 (By
CC similarity). Interacts in vitro with the basal transcription
CC factors GTF2A and TBP, suggesting an autonomous transactivation
CC function. Interacts with NCOA1, CRSP1, RBM14, the histone
CC acetyltransferases NP300 and CREBBP, and with the
CC methyltransferases EPOA6IP and HRMTLL1/PRMT2. Belongs to the
CC ASC-2/NCOA6 complex (ASCOM), which contains ASC-2/NCOA6, the
CC retinoblastoma-binding protein RBQ-3/ RBBP5, alpha- and beta-
CC tubulins, the trithorax group proteins MLL2 and MLL3, and
CC ASH2/ASCL2.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: Ubiquitous. Highly expressed in brain,
CC prostate, testis and ovary; weakly expressed in lung, thymus and
CC

small intestine.
-1- DOMAIN: Contains two Leu-Xaa-Xaa-Leu-Leu (LXXLL) motifs. Only
motif 1 is essential for the association with nuclear receptors,
while adjacent Ser-884 displays selectivity for nuclear receptors.
-1- PFM: Phosphorylated by PRKDC.
-1- PFM: Phosphorylation on Ser-884 leads to a strong decrease in
binding to ESR1 and ESR2.
-1- MISCELLANEOUS: Frequently amplified or overexpressed in colon,
breast and lung cancers.
-1- CAUTION: Ref.1 (AAFL6403) sequence differs from that shown due to
a frameshift in position 88.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
EMBL; AF177388; AAF13595.1; -
EMBL; AF208227; AAF16403.1; ALT_FRAME.
EMBL; AF245115; AAF78480.1; -
EMBL; AF128458; AAF37003.1; -
EMBL; AF171667; AAF71829.1; -
EMBL; D80003; BAA11498.2; ALT_INIT.
EMBL; AL109824; CAB92721.1; -
Genew; HGNC:15936; NCOA6.
MIM; 605299; -
GO; GO:0005634; C:nucleus; IDA.
GO; GO:0005667; C:transcription factor complex; TAS.
GO; GO:0003682; F:chromatin binding; ISS.
GO; GO:0003031; F:estrogen receptor binding; TAS.
GO; GO:0046965; F:retinoid X receptor binding; TAS.
GO; GO:0046966; F:thyroid hormone receptor binding; IDA.
GO; GO:0003713; F:transcription co-activator activity; IDA.
GO; GO:0016563; F:transcriptional activator activity; TAS.
GO; GO:0007420; P:brain development; ISS.
GO; GO:0001701; P:embryonic development (sensu Mammalia); ISS.
GO; GO:0007507; P:heart development; ISS.
GO; GO:0030099; P:myeloid blood cell differentiation; IDA.
Alignment Scores:
Pred. No.: 0.722 Length: 2063
Score: 107.50 Matches: 97
Percent Similarity: 34.70% Conservative: 47
Best Local Similarity: 23.37% Mismatches: 130
Query Match: 1.34% Indels: 141
DB: Gaps: 23
US-10-776-827-10 (1-4460) x NCO6_HUMAN (1-2063)
QY 2683 CCTAAGCACTCAGGCTCCACATC-----ATCAAC 2712
Db 1043 ProLysSerValArgLeuProValSerGlnAsnValHisProProArgGlyProLeuAsn 1062
QY 2713 CCCTTTTACCACAGAGA-----AAGAGCACTCTGGTTCTCTATCCCTTGTGCATACAG 2766
Db 1063 ProAspSerGlnArgMetProMetGlnGlnSerGlySerVal---ProValMetVal--- 1080
QY 2767 AGTTTGTATGGGGCCCTCTGGCTGTGGCTTCCATACATACAGATAACTTGCATCTGCCT 2826
Db 1081 SerLeuGlnGlyProAlaSerValProSerProAspLysGlnArgMetProMetPro 1100
QY 2827 GCACCAACACCA-----GGGATGTGGAGACATCTCCCAAC 2865
Db 1101 ValAsnThrProLeuGlySerAsnSerArgLysMetValTyrGlnGluSerProGlnAsn 1120
QY 2866 TGGCACTGTCTACCA-----GGACAGCTGCCCTTCT---GTCTCCACCTCTCAGTCC 2916
Db 1121 ProSerSerProLeuAlaGluMetAlaSerLeuProGluAlaSerGlySerGluAla 1140
QY 2917 CCCTAGATGATGGCTGGGAGAGGTGGAGGCTGACAGCTGAGACGTAGTGTGATAT 2976

```

Db 1141 Pro-----1141
QY 2977 GATCTAGGAGCGGATCAGCGGATCCGGGACCATCAAGTAAATGATGTTTCATGGCA 3036
Db 1142 -----SerValProGlyGlyProAsnAsnMetProSerHisVal 1154
QY 3037 ACTGCTTGCTCGTTGAATTAAGACAGCAGTCAGTGTCTATGCCATGACAAAGCCCTTA 3096
Db 1155 Val-----LeuProGlnAsnGlnLeuMet-----MetThrGlyPro--- 1166
QY 3097 TCTCCAGGCACAATGTCCTGCTGTCTCTTAATCAATGAGCTGTCTCACCCCGGGA 3156
Db 1167 -----LysProGlyProSerProLeuSerAlaThrGlnGly 1178
QY 3157 TGAACACCCAGAACTCACTTCTCAGTCACTTCCACAG-----3195
Db 1179 AlaThrProGlnGlnProProValAsnSerLeuProSerHisGlyHisPhePro 1198
QY 3196 -----CCGATGACTCAGAGAGGCAACCCAGAAATGGGCGCTCTTTTC 3240
Db 1199 AsnValAlaAlaProThrGlnThrSerArgProLysThrProAsnArgAlaSerProArg 1218
QY 3241 CCC---ATCACAGACTCCCTGACAACTTCTCTGGCGTAACCTAGAGAGTCCACGTGCA 3297
Db 1219 ProTyrTyrProGlnThrProAsnAsnArg-ProProSerThr-GluProSerGluLeu 1238
QY 3298 GGATAGGCCCTAAACGTTTGTGTTAATAACAGGTGCATGAAGGAGCCTAAGGCCATTG 3357
Db 1238 erLeuSerProGluArg-----1243
QY 3358 TTGATATCCACTCTCTTCTTCCACTTCTCTCATCTTTTCTCCATGTTTATGCTTC 3417
Db 1244 -----LeuAsnAlaSerIleAlaGlyLeuPhePro-----1253
QY 3418 TCTGATTCCTCTCTGCTGACAGACCCAGCCCTTATTTCTCTCCATTTT- 3476
Db 1254 -----ProGlnIleAsnIleProLeuProProArgProAsnLeuAsnArgGlyPheA 1271
QY 3477 -----CACTCTTCCAGCTCTGCTCCCTG-----AACTGCC 3507
Db 1271 spGlnGlnGlyLeuAsnProThrThrLeuLysAlaIleGlyGlnAlaProSerAsnLeu 1291
QY 3508 ACTGCAACCCATGGGA-----CTCAGGACCCAGAGCTG-----CTTGACT 3549
Db 1291 hrMetAsnProSerAsnPheAlaThrProGlnThrHisLysLeuAspSerValValAla 1311
QY 3550 CATCTGGGAGGCTAAGTTACGGGGGACAAATAATGATTCTTAAGAAGAGGCTCTCT 3609
Db 1311 snSerGlyLysGlnSerAsnSerGlyAlaThrLysArgAlaSerProSerAsnSerArgA 1331
QY 3610 AGACGACACAGGC---TCCAGAAAGACATCCCTAGCGCTGGACTTCTGAGAGCTTTA 3666
Db 1331 rgSerSerProGlySerSerArgLysThrThrProSerProGlyArgGlnAsn-----S 1349
QY 3667 GCCAGGCTCCGAC-----GGCAGCCAGAGG-----3692
Db 1349 erLysAlaProLysLeuThrLeuAlaSerGlnThrAsnAlaLeuLeuGlnAsnValG 1369
QY 3693 -----AGCGCTTCCCATGCTCTCTTCCCATGCT 3725
Db 1369 luLeuProArgAsnValLeuValSerProThrProLeuAla 1382

RESULT 11
V70K_TYMVA
ID V70K_TYMVA STANDARD; PRT; 628 AA.
AC P20131;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-FEB-1991 (Rel. 17, Last annotation update)
DE 69 kDa protein.
OS Turnip yellow mosaic virus (Australian isolate).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tymoviridae;

```

```

OC Tymovirus.
OX NCBI_TaxID=12155;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90021184; PubMed=2800335;
RA Keese P., Mackenzie A., Gibbs A.;
RT "Nucleotide sequence of the genome of an Australian isolate of turnip
RL yellow mosaic tymovirus.",
RL Virology 172:536-546(1989).
CC -!- FUNCTION: Not known.
CC -!- SIMILARITY: TO 65 TO 70 kDa PROTEIN FROM OTHER TYMOVIRUSES.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; J04373; AAA46591.1; -.
DR PIR; JQ0110; JQ0110.
DR InterPro; IPR004935; Tymo_45_70kDa.
DR Pfam; PF03251; Tymo_45kd_70kd; 1.
SQ SEQUENCE 628 AA; 68740 MW; 67CD342A09161D96 CRC64;

```

Alignment Scores:

```

Pred. No.: 0.565 Length: 628
Score: 107.00 Matches: 96
Percent Similarity: 30.57% Conservative: 37
Best Local Similarity: 22.07% Mismatches: 155
Query Match: 1.33% Indels: 147
DB: 1 Gaps: 17

```

US-10-776-827-10 (1-4460) x V70K_TYMVA (1-628)

```

QY 2748 ATCCCTTGTGTACATAGAGATTTGTCTATGGGCGCT-----CTGGCTGTGCCCTTC 2798
Db 40 LeuProMetValHisSerGluGlyThrSerAlaProThrGlnLeuLeuArgHisProAsn 59
QY 2799 ACATAACAGATAACTTCCCATCTCCCTGCCAACCAACCCAGGAGTGTGAACATCTCC 2858
Db 60 IleTrpPheGlyAsnIleProProProArgProGlnAspAsnArgAspPheSer 79
QY 2859 CCACAACTGCCACTG---CTCACGAGGACAAAGCTGCTCTCTCTCCCTCCTCAGTC 2915
Db 80 ProLeuHisProLeuValPhe-ProGlyHisHisSer-----91
QY 2916 CCCCTAGNATGATGGCTGGGGAGAGGTGGAGGTGACAGCTGAGACGCTAGTGTGAGATA 2975
Db 92 -----GlnLeuArg-----94
QY 2976 TGATCTAGGAGGCGGATCACCGGGATCCGGGACCATACAGTAACATGGTTTCATGGC 3035
Db 95 -----HisValHisGluThrGlnGlnValG1 103
QY 3036 AACTGCTTGTCTGTTGAATTA-----3057
Db 103 nGlnThrCysProGlyGluLeuLysLeuSerGlyLeuGluLeuProAlaProG1 123
QY 3058 -AGACAGAGTCACTTGTCTATTTGTCATGACAAAGCCCTTATCT---CCAGGCACAATGTC 3113
Db 123 nArgGlnHisSerLeuProLeuHisIleThrArgProSerArgPheProHisPheHi 143
QY 3114 CTTGCTGTCTCTAATCAATGAGTGTCTCTACCCAGGAGTGAACACCCAGAACT 3173
Db 143 sAlaArgArgProAspValLeuProSerLeuProAspHisGly-----ProVa 159
QY 3174 CACTTCTCAGTCACTTCCACAG---CCGATGACTCAGAAGAGCCAAACCCAGATGGGC 3230
Db 159 lLeuAlaGluThrLysProArgThrSerValArgGlnProArgSerThrThrArgGlyPr 179
QY 3231 CTCTCTTTTCCCATCACA-----GACTCCCTCGACACCT 3266

```


RP X-RAY CRYSTALLOGRAPHY (1.90 ANGSTROMS) OF 2034-2049 IN COMPLEX WITH
RP AXIN.

RP MEDLINE=20271867; PubMed=10811610;
RA Spink K.E., Polakis P., Weis W.I.;
RA "Structural basis of the axin-adenomatous polyposis coli
RT interaction.";
RL EMBO J. 19:2270-2279 (2000).

RP REVIEW ON VARIANTS.

RP MEDLINE=94154728; PubMed=8111410;
RA Nagase H., Nakamura Y.;
RT "Mutations of the APC (adenomatous polyposis coli) gene.";
RL Hum. Mutat. 2:425-434 (1993).

RP VARIANTS FAP.

RP MEDLINE=91335211; PubMed=1651563;
RA Nishisho I., Nakamura Y., Miyoshi Y., Miki Y., Ando H., Horii A.,
RA Koyama K., Utsunomiya J., Baba S., Hedge P., Markham A., Krush A.J.,
RA Petersen G.M., Hamilton S.R., Nilbert M.C., Levy D.B., Bryan T.M.,
RA Preisinger A.C., Smith K.J., Su L.-K., Kinzler K.W., Vogelstein B.;
RT "Mutations of chromosome 5q21 genes in FAP and colorectal cancer
RT patients.";
RL Science 253:665-669 (1991).

RP VARIANTS FAP.

RP MEDLINE=93265030; PubMed=1338904;
RA Miyoshi Y., Nagase H., Ando H., Ichii S., Nakatsuru S., Aoki T.,
RA Miki Y., Mori T., Nakamura Y.;
RT "Somatic mutations of the APC gene in colorectal tumors: mutation
RT cluster region in the APC gene.";
RL Hum. Mol. Genet. 1:229-233 (1992).

RP VARIANTS FAP.

RP MEDLINE=93244793; PubMed=1338691;
RA Nakatsuru S., Yanagisawa A., Ichii S., Tahara E., Kato Y.,
RA Nakamura Y., Horii A.;
RT "Somatic mutation of the APC gene in gastric cancer: frequent
RT mutations in very well differentiated adenocarcinoma and signet-ring
RT cell carcinoma.";
RL Hum. Mol. Genet. 1:559-563 (1992).

RP VARIANTS FAP TRP-1348, AND VARIANTS ASP-1118; MET-1292; VAL-1304 AND

RP SER-2502.

RP MEDLINE=93250848; PubMed=1338764;
RA Nagase H., Miyoshi Y., Horii A., Aoki T., Petersen G.M.,
RA Vogelstein B., Maher E., Ogawa M., Maruyama M., Utsunomiya J.,
RA Baba S., Nakamura Y.;
RT "Screening for germ-line mutations in familial adenomatous polyposis
RT patients: 61 new patients and a summary of 150 unrelated patients.";
RL Hum. Mutat. 1:467-473 (1992).

RP VARIANTS FAP TRP-99.

RP TISSUE=Peripheral blood lymphocytes;
RX MEDLINE=95134544; PubMed=783149;
RA Dobbie Z., Spycher M., Huerliman R., Ammann R., Roth J.,
RA Mueller A., Mueller H., Scott R.J.;
RT "Mutational analysis of the first 14 exons of the adenomatous
RT polyposis coli (APC) gene.";
RL Eur. J. Cancer 30A:1709-1713 (1994).

RP VARIANTS FAP GLY-722.

RP MEDLINE=95135430; PubMed=7833931;
RA Stella A., Montero M., Resta N., Marchese C., Susca F., Gentile M.,
RA Romio L., Pilia S., Prete F., Mareni C., Guanti G.;
RT "Four novel mutations of the APC (adenomatous polyposis coli) gene in
RT FAP patients.";
RL Hum. Mol. Genet. 3:1687-1688 (1994).

RP ERRATUM.

RA Stella A., Montero M., Resta N., Marchese C., Susca F., Gentile M.,
RA Romio L., Pilia S., Prete F., Mareni C., Guanti G.;
RL Hum. Mol. Genet. 3:1918-1918 (1994).

RP VARIANTS FAP ILE-171.
RX MEDLINE=97144176; PubMed=8990002;
RA van der Luijt R.B., Meera Khan P., Vassen H.F.A., Tops C.M.J.,
RA van Leeuwen-Cornelisse I.S.J., Wijnen J.T., van der Klift H.M.,
RA Plug R.J., Griffioen G., Rodde R.;
RT "Molecular analysis of the APC gene in 105 Dutch kindreds with
RT familial adenomatous polyposis: 67 germline mutations identified by
RT DGGE, PTT, and southern analysis.";
RL Hum. Mutat. 9:7-16 (1997).

RP VARIANTS COLORECTAL CARCINOMA THR-880; ILE-890 AND VAL-1508.

RP MEDLINE=98080146; PubMed=9419979;
RA Miyaki M., Nishio J., Konishi M., Kikuchi-Yanoshita R., Tanaka K.,
RA Muraoka M., Nagato M., Chong J.-M., Koike M., Terada T., Kawahara Y.,
RA Fukutome A., Tomiyama J., Chuganji Y., Momoi M., Utsunomiya J.;
RT "Drastic genetic instability of tumors and normal tissues in Turcot
RT syndrome.";
RL Oncogene 15:2877-2881 (1997).

RP VARIANTS LYS-1307.

RP MEDLINE=98400248; PubMed=9731522;
RA Redston M., Nathanson K.L., Yuan Z.Q., Neuhausen S.L., Satagopan J.,
RA Wong N., Yang D., Nafa D., Abrahamson J., Oczelik H.,
RA Antin-Ozarkis D., Andrulis I., Daly M., Pinsky L., Schrag D.,
RA Gallinger S., Kaback M., King M.-C., Woodage T., Brody L.C.,
RA Godwin A., Warner E., Weber B., Foulkes W., Offit K.;
RT "The APC 11307K allele and breast cancer risk.";
RL Nat. Genet. 20:13-14 (1998).

RP VARIANTS LYS-1307 AND GLN-1317.

RP TISSUE=Peripheral blood;
RX MEDLINE=98393712; PubMed=9724771;
RA Frayling I.M., Beck N.E., Ilyas M., Dove-Edwin I., Goodman P.,
RA Pack K., Bell J.A., Williams C.B., Hodgson S.V., Thomas H.J.W.,
RA Talbot I.C., Bodmer W.F., Tomlinson I.P.M.;
RT "The APC variants 11307K and E1317Q are associated with colorectal
RT tumors, but not always with a family history.";
RL proc. Natl. Acad. Sci. U.S.A. 95:10722-10727 (1998).

RP VARIANTS LYS-1307.

RP MEDLINE=98400259; PubMed=9731533;
RA Woodage T., King S.M., Wacholder S., Hartge P., Struwing J.P.,
RA McAdams M., Laken S.J., Tucker M.A., Brody L.C.;
RT "The APC 11307K allele and cancer risk in a community-based study of
RT Ashkenazi Jews.";
RL Nat. Genet. 20:62-65 (1998).

RP Alignment Scores:

Pred. No.:	0.97	Length:	2843
Score:	106.50	Matches:	75
Percent Similarity:	35.41%	Conservative:	33
Best Local Similarity:	24.59%	Mismatches:	105
Query Match:	1.33%	Indels:	92
DB:	1	Gaps:	14

US-10-776-827-10 (1-4460) x APC_HUMAN (1-2843)

QY 2797 TCACATAACAGATAACTTCCCATCTGCTGCACCAACCCAGGATGTGGAGACATCT 2856

Db 2225 SerArgGlyArgThrMetIleHisIleProGlyValArgAsnSerSerSerThrSer 2244

QY 2857 CCCACAACTGCCACTGTCTCACGAGCAAGCTGCCCTTCTCTCCACTCTCAGTCC 2916

Db 2245 ProValSerLysLysGlyProPro-----LeuLysThrProAlaSerLysSer 2260

QY 2917 CCCTAGATGATGCTGGGAGAGGTGGAGGCTGACAGCTGACAGCTAGTGTGAGATAT 2976

Db 2261 ProSerGluGlyGlnThrAlaThrThrSerProArgGlyAlaLysProSerValLysSer 2280

QY 2977 GATCTA-----GGAGGCGGATC----- 2994

Db 2281 GluLeuSerProValAlaArgGlnThrSerGlnIleGlyGlySerSerLysAlaProSer 2300


```

QY 2995 ---ACGGGATCCGGACCATCAAGTAACATGTTTTCATGCAACTGCTGCTCGTTT 3051
Db      :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
QY 2301 ArgSerGlySerArgAspSerThrProSerArgProAla 2313
Db      :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
QY 3052 GAATTAAGACAGAGTCAGTTGTCATGTCATGCCATGACAAAGCCT 3099
Db      :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
QY 2314 -----GlnGlnProLeuSerArgProGlnSerProGlyArgAsnSerIleSer 2330
Db      :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
QY 3100 CCAGGCACAAATCCCTGCTGCTCTCTTAATCAATGAGTTCGTC 3147
Db      :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
QY 2331 ProGlyArgAsnGlySerProProAsnLysLeuSerGlnLeuProArgThrSerSer 2350
Db      :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
QY 3148 CCCAGGGATGAAACACACCCAGAACTCACTTCT 3180
Db      :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
QY 2351 ProSerThrAlaSerThrLysSerSerGlySerGlyLysMetSerTyrThrSerProGly 2370
Db      :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
QY 3181 ---CAGTCACCTCCACAGCCGATGACTCAGAGAGCCAAACCCAGAAATGGGCGCTCTCT 2327
Db      :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
QY 2371 ArgGlnMetSerGlnGlnAsnLeuThrLysGlnThrGlyLeuSerLysAsnAlaSerSer 2390
Db      :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
QY 3238 TTCCCATCACAGACTCCCTGCACACCTTCTGCGGTAACTAGAGGAGTCCCAAGTGA 3297
Db      :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
QY 2391 IlePro 2397
Db      :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
QY 3298 GGATAGGCGCTAAACGTTTGT 3336
Db      :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
QY 2398 SerLysGlyLeuAsnGlnMetAsnAsnGlyAsnGlyAlaAsnLysValGluLeuSer 2417
Db      :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
QY 3337 -----GAAAGAGCGCTAAGGCCATGTTGT 3360
Db      :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
QY 2418 ArgMetSerSerThrLysSerSerGlySerGluSerAspArgSerGluArgProValLeu 2437
Db      :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
QY 3361 ATA-----TCCACCTCTCTTCCCTCACTCTCTCTCACTCTTTTC 3400
Db      :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
QY 2438 ValArgGlnSerThrPheIleLysGluAlaProSerProThr-LeuArgArgLysLeuG1 2457
Db      :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
QY 3401 ---TCCATGTTTATGCTCTCTCTGATCCCTCTCTGCTGCACACAGACAGCC---CC 3453
Db      :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
QY 2457 uGluSerAlaSerPheGluSerLeuSerProSerSer 2474
Db      :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
QY 3454 AGCCCTTTATTCCTCCATTTTTCATCTCTCCAGCCTCTGTCCTGAACTGCCATGGC 3513
Db      :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
QY 2474 oThrArgSerGlnAlaGlnThrProValLeuSerProSerLeuProAspMetSerLeuSe 2494
Db      :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
QY 3514 AACCATGGGACC 3526
Db      :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
QY 2494 rThrHisSerSer 2498
Db      :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

```

```

RESULT 14
MUC2 HUMAN STANDARD; PRT: 5179 AA.
AC Q02817; Q14878;
DT 01-JUN-1994 (Rel. 29, Created)
DE 28-FEB-2003 (Rel. 35, Last sequence update)
DE Mucin 2 precursor (Intestinal mucin 2).
GN MUC2 OR SMUC
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Intestine;
RX MEDLINE=94132002; PubMed=8300571;
RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;
RT "Molecular cloning of human intestinal mucin (MUC2) cDNA.
RT Identification of the amino terminus and overall sequence similarity
RT to prepro-von Willebrand factor.";
RL J. Biol. Chem. 269:2440-2446(1994).
RN [2]
RP SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.
RC TISSUE=Colon;

```

```

RX MEDLINE=93016075; PubMed=1400449;
RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Rothe E.-M., Lagace R.E.,
RA Kim Y.S.;
RT "The human MUC2 intestinal mucin has cysteine-rich subdomains located
RT both upstream and downstream of its central repetitive region.";
RL J. Biol. Chem. 267:21375-21383(1992).
RN [3]
RP SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.
RX MEDLINE=91358717; PubMed=1885763;
RA Toribara N.W., Gum J.R. Jr., Culhane P.J., Lagace R.E., Hicks J.W.,
RA Petersen G.M., Kim Y.S.;
RT "MUC-2 human small intestinal mucin gene structure. Repeated arrays
RT and polymorphism.";
RL J. Clin. Invest. 88:1005-1013(1991).
CC -!- FUNCTION: Coats the epithelia of the intestines, airways, and
CC other mucus membrane-containing organs. Thought to provide a
CC protective, lubricating barrier against particles and infectious
CC agents at mucosal surfaces.
CC -!- SUBUNIT: Multimeric.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: COLON, SMALL INTESTINE, COLONIC TUMORS,
CC BRONCHUS, CERVIX AND GALL BLADDER.
CC -!- PTM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR
CC INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).
CC -!- POLYMORPHISM: The number of repeats is highly polymorphic and
CC varies among different alleles.
CC -!- SIMILARITY: Contains 1 C-terminal cysteine knot-like (CTCK) domain.
CC -!- SIMILARITY: Contains 1 TIL (Trypsin inhibitory-like) domain.
CC -!- SIMILARITY: Contains 2 VWGC domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L21998; AAB95295.1; -
DR EMBL; M74027; AAB59875.1; -
DR EMBL; M94131; AAB59163.1; -
DR EMBL; M94132; AAB59164.1; -
DR PIR; A49963; A49932.
DR Genew; HGNC:7512; MUC2.
DR MIM; 158370; -
DR GO; GO:0005803; C:secretory vesicle; TAS.
DR InterPro; IPR006208; Cys_knot.
DR InterPro; IPR006207; Cys_knot_C.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR002919; TIL_Cysrich.
DR InterPro; IPR001007; VWF_C.
DR InterPro; IPR001846; VWF_D.
DR Pfam; PF00007; Cys_knot; 1.
DR Pfam; PF01826; TIL; 1.
DR Pfam; PF00093; vwc; 1.
DR Pfam; PF00094; vwd; 4.
DR SMART; SM00214; VWGC; 2.
DR SMART; SM00216; VWD; 4.
DR PROSITE; PS01185; CTCK_1; 1.
DR PROSITE; PS01225; CTCK_2; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR PROSITE; PS01208; VWFC_1; 2.
DR PROSITE; PS0184; VWFC_2; 2.
DR Glycoprotein; Repeat; Signal.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 5179 MUCIN 2.
FT DOMAIN 1401 1747 APPROXIMATE REPEATS.
FT REPEAT 1401 1416 1.
FT REPEAT 1417 1432 2.
FT REPEAT 1433 1448 3.
FT REPEAT 1449 1464 4.
FT REPEAT 1465 1471 5.
FT REPEAT 1472 1478 6.

```

[illegible]

RA Gallahan D., Callahan R.;
RT "The mouse mammary tumor associated gene INT3 is a unique member of
RT the NOTCH gene family (NOTCH4).";
RL Oncogene 14:1893-1890(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung, and Testis;
RX MEDLINE=96281668; PubMed=8681805;
RA Uyttendaele H., Marazzi G., Wu G., Yan Q., Sassoon D., Kitajewski J.;
RT "Notch4/int-3, a mammary proto-oncogene, is an endothelial
RT cell-specific mammalian Notch gene.";
RL Development 122:2251-2259(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC Rowen L., Mahairas G., Qin S., Ahearn M.E., Dankers C., Lasky S.,
RA Loretz C., Schmidt S., Tipton S., Traicoff R., Zackrone K., Hood L.;
RT "Sequence of the mouse major histocompatibility locus class III
RT region.";
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 1436-1600 FROM N.A.
RX MEDLINE=99252212; PubMed=10233982;
RA Lee J.-S., Haruna T., Ishimoto A., Honjo T., Yanagawa S.-I.;
RT "Intracisternal type A particle-mediated activation of the Notch4/int3
RT gene in a mouse mammary tumor: Generation of truncated Notch4/int3
RT mRNAs by retroviral splicing events.";
RL J. Virol. 73:5166-5171(1999).
RN [6]
RP FUNCTION.
RX MEDLINE=21244657; PubMed=11344305;
RA Uyttendaele H., Ho J., Rossant J., Kitajewski J.;
RT "Vascular patterning defects associated with expression of activated
RT Notch4 in embryonic endothelium.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:5643-5648(2001).
RN [7]
RP SEQUENCE OF 1463-1964, POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS
RP OF VAL-1463.
RX MEDLINE=21523956; PubMed=11518718;
RA Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;
RT "Murine notch homologs (N1-4) undergo presenilin-dependent
RT proteolysis.";
RL J. Biol. Chem. 276:40268-40273(2001).
RN [8]
RP POST-TRANSLATIONAL PROCESSING.
RX MEDLINE=21374376; PubMed=11459941;
RA Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;
RT "Conservation of the biochemical mechanisms of signal transduction
RT among mammalian Notch family members.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).
CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands
CC Jagged1, Jagged2 and Delta to regulate cell-fate determination.
CC Upon ligand activation through the released notch intracellular
CC domain (NICD) it forms a transcriptional activator complex with
CC RBP-J kappa and activates genes of the enhancer of split locus.
CC Affects the implementation of differentiation, proliferation and
CC apoptotic programs (By similarity). May regulate branching
CC morphogenesis in the developing vascular system.
CC -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
CC terminal fragment N(EC) which are probably linked by disulfide
CC bonds.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
CC proteolytic processing NICD is translocated to the nucleus.
CC -!- TISSUE SPECIFICITY: Highly expressed in lung, moderately in heart
CC kidney, and at lower levels in the ovary and skeletal muscle. A
CC very low expression is seen in the brain, intestine, liver and
CC testis.
CC -!- DEVELOPMENTAL STAGE: Highly expressed in endothelial cells during
CC embryonic development from 9.0 dpc.
CC -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form
CC which is proteolytically cleaved by a furin-like convertase in the
CC trans-Golgi network before it reaches the plasma membrane to yield
CC an active, ligand-accessible form. Cleavage results in a C-
CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following

CC ligand binding, it is cleaved by TNF-alpha converting enzyme
CC (TACE) to yield a membrane-associated intermediate fragment called
CC notch extracellular truncation (NEXRT). This fragment is then
CC cleaved by presenilin dependent gamma-secretase to release a
CC notch-derived peptide containing the intracellular domain (NICD)
CC from the membrane.
CC -!- PTM: Phosphorylated.
CC -!- DISEASE: Loss of the extracellular domain causes constitutive
CC activation of the Notch protein, which leads to hyperproliferation
CC of glandular epithelial tissues and development of mammary
CC carcinomas.
CC -!- SIMILARITY: Belongs to the NOTCH family.
CC -!- SIMILARITY: Contains 29 EGF-like domains.
CC -!- SIMILARITY: Contains 3 Lin/Notch repeats.
CC -!- SIMILARITY: Contains 5 ANK repeats.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M80456; AAB38377.1; -
CC EMBL; U43691; AAC52630.1; -
CC EMBL; U43691; AAC52631.1; -
CC EMBL; AF030001; AAB82004.1; -
CC EMBL; AB016771; BAA32281.1; ALT SEQ.
CC EMBL; AB016772; BAA32283.1; ALT INIT.
CC EMBL; AB016773; BAA32284.1; ALT INIT.
CC EMBL; AB016774; BAA32285.1; -
CC PIR; A38072; TMVMT3.
CC PIR; T09059; T09059.
CC HSSP; P08709; 1BF9.
CC MGD; MGI:107471; Notch4.
CC InterPro; IPR002110; ANK.
CC InterPro; IPR000152; Asx_hydroxyl_s.
CC InterPro; IPR000742; EGF 2.
CC InterPro; IPR001881; EGF Ca.
CC InterPro; IPR001438; EGF II.
CC InterPro; IPR006209; EGF-like
CC InterPro; IPR002049; Laminin_EGF.
CC InterPro; IPR008297; Notch.
CC InterPro; IPR008800; Notch_dom.
CC Pfam; PF00023; ank; 6.
CC Pfam; PF00008; EGF; 27.
CC Pfam; PF00066; notch; 2.
CC PIRSF; PIRSF002279; Notch; 1.
CC PRINTS; PR00010; EGFBLD.
CC PRINTS; PR00011; EGFBLD.
CC PRINTS; PR01452; NOTCH.
CC SMART; SM00248; ANK; 6.
CC SMART; SM00179; EGF CA; 11.
CC SMART; SM00004; NL; 2.
CC PROSITE; PS50297; ANK_REPEAT_REGION; 1.
CC PROSITE; PS50088; ANK_REPEAT; 5.
CC PROSITE; PS00010; ASX_HYDROXYL; 11.
CC PROSITE; PS00022; EGF_1; 28.
CC PROSITE; PS01186; EGF_2; 21.
CC PROSITE; PS50026; EGF_3; 27.
CC PROSITE; PS01187; EGF_CA; 9.
CC PROSITE; PS01188; EGF_3; 9.
CC KW Receptor; Transcription regulation; Activator; Differentiation;
CC Developmental protein; Repeat; ANK repeat; EGF-like domain;
CC Transmembrane; Glycoprotein; Signal; Phosphorylation; Proto-oncogene.
CC SIGNAL 1 20
CC CHAIN 21 1964
CC CHAIN 1411 1964
CC CHAIN 1428 1964
CC CHAIN 1463 1964
CC CHAIN 21 1443
CC CHAIN 1444 1464
CC CHAIN 1465 1964
CC NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 4.
CC TRANSFORMING PROTEIN INT-3.
CC NOTCH EXTRACELLULAR TRUNCATION.
CC NOTCH INTRACELLULAR DOMAIN.
CC EXTRACELLULAR (POTENTIAL).
CC POTENTIAL.
CC CYTOPLASMIC (POTENTIAL).

QY	661	CAT	-----CCTTCTTCTTCTTCTTCTTCTTCTTCTTCTCTG---	626
Db	1138	HisAsnGlyThrCysThrGluThrProGlyLeuGlyAsnProGlyPheGlnCysThrCys	 	1157
QY	625	-----GCTGTGGCCCTCTTGC-----	 	596
Db	1158	ProProAspSerProGlyProArgCysGlnArgProGlyAlaSerGlyCysGluGlyArg	 	1177
QY	595	AGCAGCAATGCAACAGATGAAGAGGAGTGATGATCACCACAGAGTGGCAGGCCACGA	 	536
Db	1178	GlyGlyAspGlyThrCysAspAlaGly-----	 	1191
QY	535	GGACTCCAGCAGATGG---TGTTGTAGATGCCCTTCTGTGTGTTTGGTCAAGATGG---	 	482
Db	1192	-----GlyAspTrpAspGlyGlyAspCysSerLeuGlyValProAspProTrpLys	 	1208
QY	481	---TGTCACAGATAGACTGAAGAGCAGTGTCTCTCATGTGCTTGGCTGGCTAGATT	 	425
Db	1209	GlyCysProHisSer-----	 	1222
QY	424	GGGCGCAGAGAACCTT-----	 	410
Db	1223	GlyArgCysHisProGlnCysAspSerGluGluCysLeuPheAspGlyTyrAspCysGlu	 	1242
QY	409	CTCGCTTACAAAGCAGGTCTCTGAGCTCAAAAGGATGCAACCTTGACCACTTGTAATTTAT	 	350
Db	1243	IleProLeuThrCysIleProAlaTyrAspGln-----	 	1254
QY	349	TCCTCTTCCAGCTTTGGGTTTGGCATGCGCACT-----	 	317
Db	1255	CysArgAspHisPheHisAsnGlyHisCysGluLysGlyCysAsnAsnAlaGluCysGly	 	1274
QY	316	---CACAGGACCTCCACACACAGCAGCAAAATGGCCAGCCAGCTTGGGAGCCTCTCTCC	 	260
Db	1275	TrpAspGlyGlyAspCysArgProGluGlyGluAspSerGlu---GlyArgProSerLeu	 	1293
QY	259	ACGTTACTTCTGTGTTTGCCT-----TGTCAGGGGATCTCGGCCAGCAACTGCT	 	212
Db	1294	AlaLeuLeuValValLeuArgProProAlaLeuAspGlnLeuLeuAlaLeuAlaArg	 	1313
QY	211	TTCTGTGAGTCAAGTCAACAGTAGTAAACAATCGGTGAACACCCGCCCTGTGTCATTAC	 	152
Db	1314	ValLeuSerLeuThrLeuArgVal-----	 	1326
QY	151	ATCTGTGATACCCGCCAGTTTCAGTGTGTCTTGGGAAGGTCACTCAGCTGCTTGAGT	 	92
Db	1327	LysAspSerGluGlyArgAsnMetValPheProTyrProGlyThrArgAlaLysGluGlu	 	1346
QY	91	CTCAGT-----	 	86
Db	1347	LeuSerGlyAlaArgAspSerSerSerTrpGluArgGlnAlaProProThrGlnProLeu	 	1366
QY	85	-----TTCTTCATCTCAAAATGGGGAACCAATCTTTCCG	 	53
Db	1367	GlyLysGluThrGluSerLeuGlyAlaGlyPheValValMetGlyValAspLeuSer	 	1386
QY	52	GTCCGCGGAAGGCCTCTCTGCAGGCCCGCAGGAATCTCTGGTTGCTGCCCC	 	2
Db	1387	-----ArgCysGlyProGluHisProAlaSerArgCysPro	 	1398

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: September 21, 2004, 15:41:06 ; Search time 461 Seconds
(without alignments)
6213.685 Million cell updates/sec

Title: US-10-776-827-10
Perfect score: 8026
Sequence: 1 cggggcagcaaccaggagat.....aattaaaggtagctagctc 4460

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1342398 seqs, 321133274 residues

Total number of hits satisfying chosen parameters: 2684796

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_P/US10776827/runat_21092004_133243_20889/app_query.fasta_1.4615
-DB=Published Applications AA -QFMT=fastan -SUFFIX=n2p.rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -SPART=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10776827@cgn_1_1_724@runat_21092004_133243_20889
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEP TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-Fgapop=6 -Fgapext=7 -Ygapop=10 -Ygapext=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA.*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description

	1	507.5	6.3	99	12	US-10-243-552-565	Sequence 565, App
2	170	2.1	2.1	88	14	US-10-106-698-6499	Sequence 6499, App
3	159	2.0	2.0	28	12	US-10-243-552-951	Sequence 951, App
4	127.5	1.6	1.6	1203	14	US-10-174-677-103	Sequence 103, App
5	127.5	1.6	1.6	1203	15	US-10-080-334-192	Sequence 192, App
6	121.5	1.5	1.5	1251	12	US-10-114-270-80	Sequence 80, Appl
7	121	1.5	1.5	314	16	US-10-437-963-136195	Sequence 136195, Sequence 13, Appl
8	117	1.5	1.5	314	16	US-10-767-701-42364	Sequence 42364, A
9	115	1.4	1.4	1747	16	US-10-437-963-168997	Sequence 168997, Sequence 13, Appl
10	113.5	1.4	1.4	451	9	US-09-965-528-13	Sequence 142, App
11	113.5	1.4	1.4	451	10	US-09-946-374-142	Sequence 282, App
12	113.5	1.4	1.4	451	12	US-10-206-915-282	Sequence 282, App
13	113.5	1.4	1.4	451	12	US-10-139-670-282	Sequence 282, App
14	113.5	1.4	1.4	451	12	US-10-201-858-282	Sequence 13, Appl
15	113.5	1.4	1.4	451	12	US-09-969-984-13	Sequence 282, App
16	113.5	1.4	1.4	451	12	US-10-205-890-282	Sequence 282, App
17	113.5	1.4	1.4	451	12	US-10-208-024-282	Sequence 282, App
18	113.5	1.4	1.4	451	12	US-10-201-853-282	Sequence 282, App
19	113.5	1.4	1.4	451	12	US-10-063-745-82	Sequence 82, Appl
20	113.5	1.4	1.4	451	12	US-10-063-512-82	Sequence 82, Appl
21	113.5	1.4	1.4	451	12	US-10-063-513-82	Sequence 82, Appl
22	113.5	1.4	1.4	451	12	US-10-063-515-82	Sequence 82, Appl
23	113.5	1.4	1.4	451	12	US-10-063-549-82	Sequence 82, Appl
24	113.5	1.4	1.4	451	12	US-10-063-569-82	Sequence 82, Appl
25	113.5	1.4	1.4	451	12	US-10-063-551-82	Sequence 82, App
26	113.5	1.4	1.4	451	12	US-10-174-581-282	Sequence 282, App
27	113.5	1.4	1.4	451	12	US-10-176-483-282	Sequence 282, App
28	113.5	1.4	1.4	451	12	US-10-176-749-282	Sequence 282, App
29	113.5	1.4	1.4	451	12	US-10-176-914-282	Sequence 282, App
30	113.5	1.4	1.4	451	12	US-10-176-915-282	Sequence 282, App
31	113.5	1.4	1.4	451	12	US-10-006-485A-142	Sequence 142, App
32	113.5	1.4	1.4	451	12	US-10-013-907A-142	Sequence 142, App
33	113.5	1.4	1.4	451	12	US-10-015-499A-142	Sequence 142, App
34	113.5	1.4	1.4	451	12	US-10-063-555-82	Sequence 82, Appl
35	113.5	1.4	1.4	451	12	US-10-063-563-82	Sequence 82, Appl
36	113.5	1.4	1.4	451	12	US-10-063-594-82	Sequence 82, Appl
37	113.5	1.4	1.4	451	12	US-10-063-554-82	Sequence 82, Appl
38	113.5	1.4	1.4	451	12	US-10-063-554-82	Sequence 82, App
39	113.5	1.4	1.4	451	12	US-10-176-484-282	Sequence 282, App
40	113.5	1.4	1.4	451	12	US-10-180-550-282	Sequence 282, App
41	113.5	1.4	1.4	451	12	US-10-183-014-282	Sequence 282, App
42	113.5	1.4	1.4	451	12	US-10-187-738-282	Sequence 282, App
43	113.5	1.4	1.4	451	12	US-10-187-740-282	Sequence 282, App
44	113.5	1.4	1.4	451	12	US-10-187-883-282	Sequence 282, App
45	113.5	1.4	1.4	451	12	US-10-194-363-282	Sequence 282, App

ALIGNMENTS

RESULT 1

US-10-243-552-565
; Sequence 565, Application US/10243552
; Publication No. US20030224379A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Zhiwei
; APPLICANT: Weng, Gezhi
; APPLICANT: Ma, Yunqing
; TITLE OF INVENTION: Novel Nucleic Acids and
; FILE REFERENCE: Polypeptides
; FILE REFERENCE: 807A
; CURRENT APPLICATION NUMBER: US/10243552
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 60/322,511
; PRIOR FILING DATE: 2001-09-13
; PRIOR APPLICATION NUMBER: PCT/US00/35017
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: PCT/US01/02623

```

; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: US 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: PCT/US01/03800
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 09/496,914
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: US 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: PCT/US01/04927
; PRIOR FILING DATE: 2001-02-26
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 998
; SOFTWARE: pt_FL_genes Version 5.0
; SEQ ID NO 565
; LENGTH: 99
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-243-552-565

Alignment Scores:
Pred. No.: 1,33e-38 Length: 99
Score: 507.50 Matches: 96
Percent Similarity: 96.97% Conservatives: 0
Best Local Similarity: 96.97% Mismatches: 3
Query Match: 6.32% Indels: 3
DB: 12 Gaps: 1

US-10-776-827-10 (1-4460) x US-10-243-552-565 (1-99)
QY 442 ATGGAGAGATCAGTGCCTTCTCAGCTCTATCTGGGACACCATCTTGACCAACACCAA 501
Db 1 MetGluArgIleSerAlaPhePheSerSerIleTrrPaspThrIleuLeuThrLysHisGln 20
QY 502 GAAGCATCTACAACACCATCTGCCTGGGAGTCTCTGGGCTGCGCTTGTGTGATC 561
Db 21 GluGlyIleTyrAsnThrIleCysLeuGlyValLeuLeuGlyLeuProLeuLeuValIle 40
QY 562 ATCACACTCTCTCATCTGTGTGCCATTGCTGTGAGCCACCGAGGAGGGCCAG 621
Db 41 IleThrLeuLeuPheIleCysHisCysTrrPaspThrProGlyLysArgGlyGln 60
QY 622 CACCCAGAG-----AGAAAGAGAGAGAGAGAGAGAGATGAAGAGACCTC 672
Db 61 GlnProGluLysAsnLysLysLysLysLysLysLysLysLysLysLysLysLysLys 80
QY 673 TGGATCTCTGCTCAACCCAGCTTCTCCAGATGGAGAGAGAGACCATCCTGCCTGTT 729
Db 81 TrpIleSerAlaGlnProLysLeuLeuGlnMetGluLysArgProSerLeuProVal 99

RESULT 2
US-10-106-698-6499
; Sequence 6499, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 6499
; LENGTH: 88
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-106-698-6499

Alignment Scores:
Pred. No.: 6,71e-06 Length: 28
Score: 159.00 Matches: 28
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.98% Indels: 0
DB: 12 Gaps: 0

US-10-776-827-10 (1-4460) x US-10-243-552-951 (1-28)

; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: US 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: PCT/US01/03800
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 09/496,914
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: US 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: PCT/US01/04927
; PRIOR FILING DATE: 2001-02-26
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 998
; SOFTWARE: pt_FL_genes Version 5.0
; SEQ ID NO 951
; LENGTH: 28
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-243-552-951

Alignment Scores:
Pred. No.: 6,71e-06 Length: 28
Score: 159.00 Matches: 28
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.98% Indels: 0
DB: 12 Gaps: 0

US-10-776-827-10 (1-4460) x US-10-106-698-6499 (1-88)
QY 2844 GTGAAGACATCTCCCAACAACTGCCTCCTCACCAGGACAAAGTGCCTTCTGTCTC 2903
Db 51 MetGluAspIleSerProGlnLeuProLeuLeuThrArgThrSerCysProSerCysLeu 70
QY 2904 CACCTCTCAGTCCCTCCCTAGATGATGGATGGCTGG-GGAGAGGTGGAGGCTGCACAGC 2956
Db 71 HisLeuSerValProLeuGluTrpMetAlaGlyGlyGluValGluAlaAspSer 88

RESULT 3
US-10-243-552-951
; Sequence 951, Application US/10243552
; Publication No. US20030224379A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Zhiwei
; APPLICANT: Weng, Gezhi
; APPLICANT: Ma, Yunqing
; TITLE OF INVENTION: Novel Nucleic Acids and
; FILE REFERENCE: 807A
; CURRENT APPLICATION NUMBER: US/10/243,552
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 60/322,511
; PRIOR FILING DATE: 2001-09-13
; PRIOR APPLICATION NUMBER: PCT/US00/35017
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: PCT/US01/02623
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: US 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: PCT/US01/03800
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 09/496,914
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: US 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: PCT/US01/04927
; PRIOR FILING DATE: 2001-02-26
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 998
; SOFTWARE: pt_FL_genes Version 5.0
; SEQ ID NO 951
; LENGTH: 28
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-243-552-951
```


Table 1. Demographic characteristics of study population

Db 941 erAlaSerProGlnPro-AlaPheHisLeuLysProAspThrProVal-----SerVal 958
Qy 898 AGGACCAC-----GTGCACAATTCTTGGATACTCTCTTGGCAG 936
Db 959 LysLysHisValIleGlnGluLeuProLeuAspAsnThrPheValGlyGly----- 976
Qy 937 CTATGTGTCCAATAGCAATGCTCTTACTGTCCAGACCAGGATCCCTCC-----CA 987
Db 977 -----CysAspThrLeuSer-LysArgSerSerThrSerSerAspHis 990
Qy 988 CCTGTCTGTGCATACCCCATCATGCAAGACCAAGACATTT-----AT 1032
Db 990 spHeSerAlaSer-----GluCysSerSerGlnGlyGlyPheHisLysThrLysGlyProLeu 1008
Qy 1033 CCATACATCTCAATATGGTTCCTCCCAAGTGTGTGCATGCACGTACAGTAAACACACACACAA 1092
Db 1008 uHisThrArgGlnSerGlnArgValThrPheHisLeuProAsp----- 1023
Qy 1093 ATTCAAGTAGCAGGTACGTGGGCAAGTATATTCGTCTCATCAATGGTCTATGGCTATGT 1152
Db 1024 -----GlySerGlnGluSerCysSerAspSerGlyLeuGlyAspHis 1037
Qy 1153 ACTTTGTGCAGGAGTACATTAATCTACATGTCACAAAATGTCT----- 1196
Db 1037 sGluProValGlySerGlyThrLeuLeuSerHis-ProLeuProLeuValGlnProGlnA 1057
Qy 1197 -----CATGGGAAGCCTTCCAGATTTCAGACACATATATA-----CAATTTC 1239
Db 1057 spGluPheThrAspGlnAlaSerProAspLysArgThrGluAlaAspGlyAsnSerAspP 1077
Qy 1240 CTAACACAGCAGGCCCCCATACACCATCTATTCCATAAACCACTCAGGTTACAGATGCAT 1299
Db 1077 roAsnSerAspGlyProLeuGlyProArg----- 1086
Qy 1300 GCTTTCTTCTTCTAATCTACATACATAACTTTTACTGAAGTACTCATATAATTGGACATT 1359
Db 1087 -----GlyLeuAlaGluAlaThrGluMetCysThrGlnGluCysLeuValLeuGlyHis 1105
Qy 1360 CCAGCAACCTGTACAGTCCCACTTGTGTCTGTCTGTATACAGACACACACCAAG 1413
Db 1105 erAspAsn-CysTrpMetPro---ProGlyLeuGlyProTyGlnHisProLys 1121

RESULT 6

US-10-114-270-80
; Sequence 80, Application US/10114270
; Publication No. US20040030110A1
; GENERAL INFORMATION:
; APPLICANT: Guo, Xiaojia
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Miller, Charles E.
; APPLICANT: Malyankar, Uriel M.
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Patturajan, Meera
; APPLICANT: Liu, Ziaohong
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Li, Li
; APPLICANT: Vernet, Corine
; APPLICANT: Zerhusen, Bryan D.
; APPLICANT: Gorman, Linda
; APPLICANT: Shenoy, Suresh G.
; APPLICANT: Pena, Carol E.A.
; APPLICANT: Smithson, Glenda
; APPLICANT: Burgess, Catherine E.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Padigaru, Muraidhara
; APPLICANT: Shimkets, Richard A.
; APPLICANT: Gangolli, Esha A.
; APPLICANT: Taupier Jr., Raymond J.
; APPLICANT: Casman, Stacie J.
; APPLICANT: Ji, Weizhen
; APPLICANT: Anderson, David W.
; APPLICANT: Liete, Mario W.

; PRIOR APPLICATION NUMBER: 60/330,307
; PRIOR FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: 60/278,796
; PRIOR FILING DATE: 2001-03-26
; PRIOR APPLICATION NUMBER: 60/281,521
; PRIOR FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: 60/276,677
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/311,595
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/270,220
; PRIOR FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: 60/274,295
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/318,526
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/286,548
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: 60/291,765
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 60/270,797
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/276,400
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/270,810
; PRIOR FILING DATE: 2001-02-23
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 192
; LENGTH: 1203
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-080-334-192

Alignment Scores:
Pred. No.: 0.0325 Length: 1203
Score: 127.50 Matches: 83
Percent Similarity: 38.12% Conservative: 55
Best Local Similarity: 22.93% Mismatches: 125
Query Match: 1.59% Indels: 99
DB: 15 Gaps: 18

US-10-776-827-10 (1-4460) x US-10-080-334-192 (1-1203)

Qy 496 CACCAAGAGGCATCTACACACCATCTGCTGGAGTCTCTCTGGCCCTGCCACTCTTG 555
Db 807 TyrGlnAsnGluAspTryLeuThrIleMetIleAlaIleIleAlaGlyAlaMetValVal 826
Qy 556 GTGATC-----ATCACACTCTCTTCTCATCTGTGTGCTATGCTGC----- 594
Db 827 IleValValIlePheValThrValLeuValArgCysArgHisAlaSerArgPheLysAla 846
Qy 595 -----TGGAGCCCAACAGGCAAGAGAGGGCCAGACGCCA 627
Db 847 AlaGlnArgSerLysGlnGlyAlaGluTrpMetSerProAsnGlnGluAsnLysGlnAsn 866
Qy 628 GAGAAGAAAAG 681
Db 867 Lys 886
Qy 682 GCTCAACCCAGCTTCTCCAGAT-----GGAGAAGAGACCA 717
Db 887 IleGlu--GluSerLysProAspAspAlaValHisGluProIleAsnGlyThrIleSerL 906
Qy 718 TCAGTCTGCTTTAGTTAGGAG 777
Db 906 euProAlaGluLeuGluGlnSerIleGlyArgPheAsp---TigGlyProAlaPro- 924
Qy 778 TCTGACCAACACACAGATTTTCAGAACCCCTGAAATATGCACTATGTCCATGTCCACACA 837
Db 925 -----ProThrThrPheLysProAsnSerProAspLeuAlaLys-----HisTyrLys 941
Qy 838 GAGTAACTACTCAACCAAGGACAAACCTCAGACTAAGTGTCCAGGTGGAGGCGAGTCCC 897

```

; APPLICANT: Rastelli, Luca
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: Stone, David J.
; APPLICANT: MacDougall, John R.
; APPLICANT: Rothenberg, Mark E.
; TITLE OF INVENTION: No. US20040030110A1el Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-322C
; CURRENT APPLICATION NUMBER: US/10/114,270
; CURRENT FILING DATE: 2002-11-27
; PRIOR APPLICATION NUMBER: 60/281,086
; PRIOR FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: 60/281,136
; PRIOR FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: 60/281,863
; PRIOR FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: 60/281,906
; PRIOR FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: 60/282,020
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: 60/282,930
; PRIOR FILING DATE: 2001-04-10
; PRIOR APPLICATION NUMBER: 60/282,934
; PRIOR FILING DATE: 2001-04-10
; PRIOR APPLICATION NUMBER: 60/283,512
; PRIOR FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 60/283,710
; PRIOR FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/284,234
; PRIOR FILING DATE: 2001-04-17
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 470
; SEQ ID NO 80
; LENGTH: 1251
; TYPE: PRT
; ORGANISM: Homo sapiens
; UN-10-114-270-80

```

Alignment Scores:					
Pred. No.:	0.122	Length:	1251		
Score:	121.50	Matches:	127		
Percent Similarity:	32.39%	Conservative:	56		
Best Local Similarity:	22.48%	Mismatches:	176		
Query Match:	1.51%	Indels:	207		
DB:	12	Gaps:	27		
US-10-776-827-10 (1-4460) x US-10-114-270-80 (1-1251)					
QY	2700	CCCACTCATCAACCCCTTGACACAGAGAAGACACTCTGGTTCTCTATCCCTTGTC	2759		
Db	746	ProThrSerProLeuProPheSerSerThrProGluPheProValProLeuSer	765		
QY	2760	CATAGAGATTGTCATGGGGCCTCTGGCTGCGCTTCACATAACAGATAACTGCCA	2819		
Db	766	GlnCysProIrpSerSerLeuProThrThrSerProProThr	781		
QY	2820	TCTGCTCTACCAAAACCCAGGGATGTGGAGACATCTCCACAACTGCCACTGCTCACC	2879		
Db	782	ProThrCysSerGln	798		
QY	2880	AGGACAAGCTGCCCTTCTGTCTCCACTCTCAGTCCCCCTTAGAATGATGGCTGGGAG	2939		
Db	798	oSerThrSerSerPhePro--SerThrAlaAlaPro--	810		
QY	2940	AGGTGGAGGCTGACAGCTGAGCGTAGCTGACATATGATCTAGAGGGCGGATCACCGG	2999		
Db	810		810		
QY	3000	GATCCGGGACCATACAAGTAACATGTTTCATGGGAAGTCTGCTGCTTGAATTAAAG	3059		
Db	811		823		
QY	3060	A-----CAGCAGTCAGTTGTCATTGCCATGACA-----	3087		

QY 3932 GTTACTTTTACATTAGTATTACCCCACTCAGACATCATCTCAGAAATGAT---CAA 3988
 Db 1156 -----ProThrSerArgArgAsnSerLeuGln 1164
 QY 3989 TGCAGGACTCTTCTGTGACACAAATGTCCTCCAGCCCTCCCT----- 4030
 Db 1165 ArgSerGluProThrCys-----ValProAlaIleProAlaThrGlnGluAlaGlu 1181
 QY 4031 ---GTCACCGCTTCGCCCATGTAGAGTCGTAGGTCTGTAGAGATGAGGAATGTGGCTGTC 4087
 Db 1182 AlaGlyGluPheLeuGluProGlyArg-----ArgArgLeuGlnAsnHisSerPro 1198
 QY 4088 TCACCCCTTG 4096
 Db 1199 SerAlaLeu 1201

RESULT 7
 US-10-437-963-136195
 ; Sequence 136195, Application US/10437963
 ; Publication No. US20040123343A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa, Thomas J.
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Zhou, Yinhua
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Wu, Wei
 ; APPLICANT: Boukharov, Andrey A.
 ; APPLICANT: Barbazuk, Brad
 ; APPLICANT: Li, Ping
 ; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
 ; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
 ; FILE REFERENCE: 38-21(53221)B
 ; CURRENT APPLICATION NUMBER: US/10/437,963
 ; CURRENT FILING DATE: 2003-05-14
 ; NUMBER OF SEQ ID NOS: 204966
 ; SEQ ID NO 136195
 ; LENGTH: 314
 ; TYPE: PRT
 ; ORGANISM: Oryza sativa
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: PAT_MRT4530_37798C.1.pep
 US-10-437-963-136195

Alignment Scores:
 Pred. No.: 0.0747 Length: 314
 Score: 121.00 Matches: 47
 Percent Similarity: 43.43% Conservative: 29
 Best Local Similarity: 26.86% Mismatches: 57
 Query Match: 1.51% Indels: 42
 DB: 1.6 Gaps: 10

US-10-776-827-10 (1-4460) x US-10-437-963-136195 (1-314)

QY 151 TCTGAATGCACAGCGGGTGTTCAGCGATTGTTTACTAGTTGAAGTGCACCTCCAGGA 210
 Db 66 SerGluLeuGlyAspGluLeuSerProCysMetCysGlyThr----- 81
 QY 211 AAGCAGTTCGCGCGAGATCC---CCTGACACCAAGCAAGCAAGTAACGTGGAAGAG 267
 Db 82 GlnGlnPheValHisArgSerCysLeuAspHis----- 95
 QY 268 GCTCCCAAGTGTGCGCCATTTGCTGTGTGTGTGGAGGTGCTGTCAAGTGCATGCC 327
 Db 96 ValLysGlu---GlyPheAlaPheSerHisCysThrThrCysLysAlaGlnPheHisLeu 114
 QY 328 CAACCCAAAGCTGCAAGGAGTAATTAACAGTGTGCAAGTGTGCAATCTTTTGAGCT 387
 Db 115 ArgValGluThrTrpGluAspAsnSerTrpArgLysMetLysPheArgIlePheValAla 134
 QY 388 CAGGACCTGCTGTGAAGCGAGAGGGTTCCTGCGCCCTAATCTAGCCAAAGCACCAGT--- 444
 Db 135 ArgAspValIleLeuVal-----PheLeuAlaValGlnLeuProSerAlaMetIle 151

QY 445 -----GAGAGAAATCAGTCCCTTC-----TTGAGCTCTATC 474
 Db 152 GlyAlaIleSerTyrPheLeuAspArgAspGlySerPheArgAsnSerPheSerAspGly 171
 QY 475 TGGGACACCATCTTGTGACCAACACCAAGAGCATCTACACACCATCTGCTGGAGTTC 534
 Db 172 TrpAspArgPheLeuSerLysHisProIleProPheTyr-----TyrCysIleGlyVal 189
 QY 535 CTCCTGGGCGCTTCGTGATCATCATCATCTCTCTCTTCATCTGT----- 582
 Db 190 ValValPhePheValLeuLeuGlyPhePheGlyLeuIleLeuHisCysSerSerPheAsn 209
 QY 583 -----TGC---CATTCCTGCTGG 597
 Db 210 AspAsnGlnAspProCysLeuAlaGlyCysArgAsnCysCysTyr 224

RESULT 8
 US-10-767-701-42364
 ; Sequence 42364, Application US/10767701
 ; Publication No. US20040172684A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Zhou, Yinhua
 ; APPLICANT: Cao, Yongwei
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
 ; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
 ; FILE REFERENCE: 38-21(53535)B
 ; CURRENT APPLICATION NUMBER: US/10/767,701
 ; CURRENT FILING DATE: 2004-01-29
 ; NUMBER OF SEQ ID NOS: 63128
 ; SEQ ID NO 42364
 ; LENGTH: 314
 ; TYPE: PRT
 ; ORGANISM: Sorghum bicolor
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: SORBI-28MAY03-C61842_1.pep
 US-10-767-701-42364

Alignment Scores:
 Pred. No.: 0.178 Length: 314
 Score: 117.00 Matches: 45
 Percent Similarity: 42.61% Conservative: 30
 Best Local Similarity: 25.57% Mismatches: 57
 Query Match: 1.46% Indels: 44
 DB: 1.6 Gaps: 10

US-10-776-827-10 (1-4460) x US-10-767-701-42364 (1-314)

QY 151 TCTGAATGCACAGCGGGTGTTCAGCGATTGTTTACTAGTTGAAGTGCACCTCCAGGA 210
 Db 66 SerGluIleGlyAspGluLeuLeuSerProCysMetCysGlyThr----- 81
 QY 211 AAGCAGTTCGCGCGAGATCC---CCTGACACCAAGCAAGCAAGTAACGTGGAAGAG 267
 Db 82 GlnGlnPheValHisArgSerCysLeuAspHis----- 95
 QY 268 GCTCCCAAGTGTGCGCCATTTGCTGTGTGTGTGGAGGTGCTGTCAAGTGCATGCC 327
 Db 96 ValLysGlu---GlyPheAlaPheSerHisCysThrThrCysLysAlaGlnPheHisLeu 114
 QY 328 CAACCCAAAGCTGGAAGAGTAATTAACAGTGTGCAAGTGTGCAATCTTTTGAGCT 387
 Db 115 ArgValGluThrTrpGluAspAsnSerTrpHisLysMetLysPheArgIlePheValAla 134
 QY 388 CAGGACCTGCTT-----GTAAGCCGAGAGGGTTCCTG 420
 Db 135 ArgAspValLeuLeuValPheLeuAlaValGlnLeuThrIleAlaIleGlyAlaIle 154
 QY 421 GCCCTAATCTAGCCAAAGCACCATGAGGAATCATAGTGCCTTC-----TTAGCTCT 471
 Db 155 AlaTyrPhe-----LeuAspArgAspGlySerPheArgAsnSerPheSerAsp 170

QY 472 ATCTGGGACACCATCTTTGACCAACACCAAGAGGCATCTACACACCATCTGCTGGGA 531
 Db 171 GlyTpsAArgPheLeuSerlybHsProIleProPheTyr-----TyrCysIleGly 188
 QY 532 GTCTCTGGGCTGCGCATCTTTGGTGATCATCACACTCTCTCTCATCTGT----- 582
 Db 189 ValValValPhePheValLeuLeuGlyPhePheGlyLeuIleValHisCysSerSerPhe 208
 QY 583 -----TGC-----CATCTGCTGG 597
 Db 209 AsnAspAsnGlnAspProCysLeuAlaGlyCysArgAsnCysCysTyr 224

RESULT 9
 US-10-437-963-168997
 ; Sequence 168997, Application US/10437963
 ; Publication No. US20040123343A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa, Thomas J.
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Zhou, Yihua
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Wu, Wei
 ; APPLICANT: Boukharov, Andrey A.
 ; APPLICANT: Barbazuk, Brad
 ; APPLICANT: Li, Ping
 ; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 ; FILE REFERENCE: 38-21(53221)B
 ; CURRENT APPLICATION NUMBER: US/10/437,963
 ; CURRENT FILING DATE: 2003-05-14
 ; NUMBER OF SEQ ID NOS: 204966
 ; SEQ ID NO 168997
 ; LENGTH: 1747
 ; TYPE: PRT
 ; ORGANISM: Oryza sativa
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: PAT_MRT4530_67458C.1.psp
 US-10-437-963-168997

Alignment Scores:
 Pred. No.: 0.58 Length: 1747
 Score: 115.00 Matches: 77
 Percent Similarity: 33.45% Conservative: 22
 Best Local Similarity: 26.01% Mismatches: 109
 Query Match: 1.43% Indels: 88
 DB: 16 Gaps: 14

US-10-776-827-10 (1-4460) x US-10-437-963-168997 (1-1747)

QY 160 ACAGGGCGGTGTTTCAGCGATTGTTTACTACGTTCAGCTGACCTCCAGGAAGCAGTTC 219
 Db 1056 ThrProGlnValValSerAlaLeuVal-ValGluArgAspGluAspAsnProHisSe 1075
 QY 220 TGGCGGA---GATCCCTGCACAAACCAAGCAAGAAAGTAACGTGGAAGAGGCTCCCAA 276
 Db 1075 rAlaHisProHisProValSerThrArgProGlyArgGluGlnGlyGluAlaProGI 1095
 QY 277 CTGGCTGGCATTTTGTCTGTGTGTGGAGGTGCT----- 313
 Db 1095 uProAsnGlyGlyProArgProProThrAlaGlyAlaGlyProProProAlaCysProTh 1115
 QY 314 -GTCAGTGGCATGCCCAACCAAGCTGGAGAGGAATAATATCAAGTGTCAAGTT 372
 Db 1115 rValProGlyAlaProAspProLyAspGlyProGlyAlaThrAla-----GlyArgPr 1133
 QY 373 GCATCTTTTGGCTCAGGACCT-----GCTTGAAGCCGAGAG 411
 Db 1133 oHis-LeuLeuProSerAspProGluValValGlyThrGluAlaGluCysAlaProArg 1153
 QY 412 GGTCTCTGGCCCTTAATCTAGCCAAGCACCATTGGAGAGATCATGTCCTTCTTCAGTCT 471
 Db 1153 lyLeuSerAspGluGlu----- 1158

QY 472 ATCTGGGACACCATCTTTGACCAACACCAAGAGGCATCTA-----CAACACCATCTGC 525
 Db 1159 -----HisProValSerThrArgProGlyArgGluGlnGlyGluAlaProGluP 1176
 QY 526 CTGGGAGTCTCTCTGGGCTGCGCATCTTTGGTGATCATCACACTCTCTCTCTCTTTC 585
 Db 1176 roAsnGlyGlyProArgProProThrAlaGlyAlaGlyProLeuProAla---CysLeu 1195
 QY 586 CATTGCTGCTGGAGCCACACAGGAGGCGGCGCAGCAGCAGGAGAGAGAGAGAGAGAG 645
 Db 1195 hrValProGlyAlaProAspProGlnAspGlyProGlyAlaThrAla----- 1210
 QY 646 AAGAAGAAGAGATGAAGAAGACCTCTCTGATCTCTGTCACACCCCAAGTCTTCCAGATG 705
 Db 1211 -----GlyArgPro-----ArgLeuSerProSerA 1219
 QY 706 GAGAAGAGACCATCACTGCTGCTTTAGTTAGGAGGAGGAGGAGGAGGAGGAGGAG 765
 Db 1219 spProGlu-----ValValGlyThrGluAlaGluCysAlaPro----- 1231
 QY 766 GCTAAGCCTCTCTCTGACCAACACACAGACATTTTCAGGAACCCCTGAAATATGCACTATG 825
 Db 1232 -----ArgGlyLeuSerAsp----- 1236
 QY 826 TCATGTCACAGAGTAATACTACTCAACCAAGAGAAACAACCTCAGACTAAGTGTCCAGTG 885
 Db 1237 --GluGluHisProValSerThrArgProGlyArgGluGlnGlyGluAlaProSerA 1256
 QY 886 GAGGCGAGTCCCGAGGACCACTGTCGACATTTCTTGATCTCTCTGAGTACTCTCTG 945
 Db 1256 rgThrAlaAlaGlnGlyProArgProGluLeu-----AlaLeuCysP 1271
 QY 946 CAATAGCAATGCTCTTACTGACAGACCCAGGATGCTCTCCACCC 989
 Db 1271 roLeuValArgCysGlnGlu-----ProProThr 1281

RESULT 10
 US-09-965-528-13
 ; Sequence 13, Application US/09965528
 ; Publication No. US20020187523A1
 ; GENERAL INFORMATION:
 ; APPLICANT: INCYTE GENOMICS, INC.
 ; APPLICANT: TANG, Y. Tom
 ; APPLICANT: YUE, Henry
 ; APPLICANT: LAL, Preeti
 ; APPLICANT: BURFORD, Neil
 ; APPLICANT: BANDMAN, Olga
 ; APPLICANT: BAUGHN, Mariah R.
 ; APPLICANT: AZIMZAI, Yalda
 ; APPLICANT: LU, Dyung Aina M.
 ; APPLICANT: PATTERSON, Chandra
 ; TITLE OF INVENTION: EXTRACELLULAR SIGNALING MOLECULES
 ; FILE REFERENCE: PF-0701 USA
 ; CURRENT APPLICATION NUMBER: US/09/965,528
 ; CURRENT FILING DATE: 2001-09-26
 ; PRIOR APPLICATION NUMBER: 60/134,949
 ; PRIOR FILING DATE: 1999-05-19
 ; PRIOR APPLICATION NUMBER: 60/144,270
 ; PRIOR FILING DATE: 1999-07-15
 ; PRIOR APPLICATION NUMBER: 60/146,700
 ; PRIOR FILING DATE: 1999-07-30
 ; PRIOR APPLICATION NUMBER: 60/157,508
 ; PRIOR FILING DATE: 1999-10-04
 ; NUMBER OF SEQ ID NOS: 55
 ; SOFTWARE: PERL Program
 ; SEQ ID NO 13
 ; LENGTH: 451
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; OTHER INFORMATION: incyte ID No. US20020187523A1 2267403CD1
 US-09-965-528-13

Alignment Scores:
Pred. No.: 0.447 Length: 451
Score: 113.50 Matches: 90
Percent Similarity: 33.16% Conservativity: 38
Best Local Similarity: 23.32% Mismatches: 124
Query Match: 1.41% Indels: 134
DB: 9 Gaps: 23

US-10-776-827-10 (1-4460) x US-09-965-528-13 (1-451)

QY 318 GTGGATCCCAACCCAGAGCTGCAAGAGGAATAAATACAAAGTGGTCAAGTTGCATC 377
DB 79 ValHisCysProGlnPro-----ValThrGluProGlnGlnCysCys 92
QY 378 CTTTGTGAGCTCAGGACCTCTGTGT---AAGCCGAGAGGTTCTCTGGC----- 422
DB 93 -----ProLysCysValGluProHisThrProSerGlyLeuArgAlaPro 107
QY 423 CCTATCTAGCAACGAC-----CATGGAGAG----- 449
DB 108 ProLysSerCysGlnHisAsnGlyThrMetTyrGlnHisGlyGluLeuPheSerAlaHis 127
QY 450 -----AATCAGTGCCTTCTT-----CAG 467
DB 128 GluLeuPheProSerArgLeuProAsnGlnCysValLeuCysSerCysThrGluGlyGln 147
QY 468 CTCATCTCTGGACACCATCTTGACCAACACCAAGAGGCATCTACACCATCTGCTCT 527
DB 148 IleTyrCysGly-----LeuThrCysProGlu-----Pro 158
QY 528 GGGAGTCTCTCTGGGCTGGCCTCTTGTGTGATCATCACACCTCTCTCTCTCTCTCT 587
DB 159 GlyCysPro-----AlaProLeuProLeuProAsp 168
QY 588 TTGCTGCTGAGCCACCAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 647
DB 169 SerCysCysGlnAlaCysIysAspGlu-----AlaSerGlnSerAspGluGlu 185
QY 648 GAAGAA-----GAAGATGAAGAGACCTCTGATCTCTGCTCAACCCAGCT 695
DB 186 AspSerValGlnSerLeuHisGlyValArgHisProGlnAspProCysSerSerAspAla 205
QY 696 TCTCCAGATGAGAGAGACATCATCTGCTGTTAGTGGAGGAGGAGGAGGAGGAGGAG 755
DB 206 Gly-Arg-----LysArgGlyProGlyThrProAlaProThrGlyLeu-SerAlaProLeu 224
QY 756 CTTTCTGAGGCTAAGCTCTCTCTGACACACACAGACATTTTCAGGAACCCCTGAATA 815
DB 224 erPhelle-----ProArgHisPheArg----- 231
QY 816 ATGCATATGTCCATGTCCAGAGTAACTACTCAACC-----AAGG 857
DB 232 -----ProLysGlyAlaGlySerThrThrValLysIleValLeuLysG 246
QY 858 AACAACTCAGACTAAGTCTCCAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 917
DB 246 lu-LysHisLysLysAlaCysValHisGlyGlyLysThrTyrSer---HisGlyGluVal 264
QY 918 TTGATATCTGCTTGGCAGCTATGTGCTCAATA---GCAATGTCTCTTACTGACACCCA 974
DB 265 TrpHisProAlaPheArgAlaPheGlyProLeuProCysIleLeuCysThrCysGluAsp 284
QY 975 GGCATGCTCTCC-----ACCTGTCTCTGGCATATCCCATATCCCAAGTGTGACAC 1022
DB 285 GlyArgGlnAspCysGlnArgValThrCysProThrGluTyrPro---CysArgHisPro 303
QY 1023 GAA-----CAT 1028
DB 304 GluLysValAlaGlyLysCysCysLysIleCysProGluAspLysAlaAspProGlyHis 323
QY 1029 TTATCCATACATCTCAATATGTTGCCAAGTGTGTCACATGTCAGGTAAACACACACAC 1088

RESULT 11

US-09-946-374-142
; Sequence 142, Application US/09946374
; Publication No. US20030073129A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gueney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William L.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: P2830P1C1
; CURRENT APPLICATION NUMBER: US/09/946,374
; CURRENT FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: 60/098716
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098723
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098749
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098750
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098803
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/098821
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/098843
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/099536
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099596
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099598
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099602
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099642
; PRIOR FILING DATE: 1998-09-09

Db 324 SerGluLeuSerSerThrArgCysProLysAlaProGlyArgValLeuValHisThrSer 343
QY 1089 ACAAATTCAGGTACAGGTACGTGGCAAGTATATTCTGTCTCATCAATATGTCATTTGGCT 1148
DB 344 ValSerProSerProAspAsnLeuArgPheAlaLeu----- 356
QY 1149 ATGTACTTTGTGAGGGAAGTACATTATCTACAGTCAAAAATGCTCTCATGGGAAGGCC 1208
DB 357 -----GluHisGluAlaSerAspLeuValGluIleTyrLeuTrpLysLeu 371
QY 1209 TTGCCAGATTCAGACACATATATATCTTCTTAACCAAGCAAGGC-----CCCATACA 1262
DB 372 ValLysAspGluGluThrGluAlaGlnArgGlyGluValProGlyProArgProHisSer 391
QY 1263 CCATCTATTTCCA 1274
DB 392 GlnAsnLeuPro 395

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525
---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

7	PRIOR FILING DATE: 1998-09-24
7	PRIOR APPLICATION NUMBER: 60/101743
7	PRIOR FILING DATE: 1998-09-24
7	PRIOR APPLICATION NUMBER: 60/101915
7	PRIOR FILING DATE: 1998-09-24
7	PRIOR APPLICATION NUMBER: 60/101916
7	PRIOR FILING DATE: 1998-09-24
7	PRIOR APPLICATION NUMBER: 60/102207
7	PRIOR FILING DATE: 1998-09-29
7	PRIOR APPLICATION NUMBER: 60/102240
7	PRIOR FILING DATE: 1998-09-29
7	PRIOR APPLICATION NUMBER: 60/102307
7	PRIOR FILING DATE: 1998-09-29
7	PRIOR APPLICATION NUMBER: 60/102330
7	PRIOR FILING DATE: 1998-09-29
7	PRIOR APPLICATION NUMBER: 60/102331
7	PRIOR FILING DATE: 1998-09-29
7	PRIOR APPLICATION NUMBER: 60/102484
7	PRIOR FILING DATE: 1998-09-30
7	PRIOR APPLICATION NUMBER: 60/102487
7	PRIOR FILING DATE: 1998-09-30
7	PRIOR APPLICATION NUMBER: 60/102570
7	PRIOR FILING DATE: 1998-09-30
7	PRIOR APPLICATION NUMBER: 60/102571
7	PRIOR FILING DATE: 1998-09-30
7	PRIOR APPLICATION NUMBER: 60/102684
7	PRIOR FILING DATE: 1998-10-01
7	PRIOR APPLICATION NUMBER: 60/102687
7	PRIOR FILING DATE: 1998-10-01
7	PRIOR APPLICATION NUMBER: 60/102965
7	PRIOR FILING DATE: 1998-10-02
7	PRIOR APPLICATION NUMBER: 60/103258
7	PRIOR FILING DATE: 1998-10-06
7	PRIOR APPLICATION NUMBER: 60/103314
7	PRIOR FILING DATE: 1998-10-07
7	PRIOR APPLICATION NUMBER: 60/103315
7	PRIOR FILING DATE: 1998-10-07
7	PRIOR APPLICATION NUMBER: 60/103328
7	PRIOR FILING DATE: 1998-10-07
7	PRIOR APPLICATION NUMBER: 60/103395
7	PRIOR FILING DATE: 1998-10-07
7	PRIOR APPLICATION NUMBER: 60/103396
7	PRIOR FILING DATE: 1998-10-07
7	PRIOR APPLICATION NUMBER: 60/103401
7	PRIOR FILING DATE: 1998-10-07
7	PRIOR APPLICATION NUMBER: 60/103449
7	PRIOR FILING DATE: 1998-10-06
7	PRIOR APPLICATION NUMBER: 60/103633
7	PRIOR FILING DATE: 1998-10-08
7	PRIOR APPLICATION NUMBER: 60/103678
7	PRIOR FILING DATE: 1998-10-08
7	PRIOR APPLICATION NUMBER: 60/103679
7	PRIOR FILING DATE: 1998-10-08
7	PRIOR APPLICATION NUMBER: 60/103711
7	PRIOR FILING DATE: 1998-10-08
7	PRIOR APPLICATION NUMBER: 60/104257
7	PRIOR FILING DATE: 1998-10-14
7	PRIOR APPLICATION NUMBER: 60/104987
7	PRIOR FILING DATE: 1998-10-20
7	PRIOR APPLICATION NUMBER: 60/105000
7	PRIOR FILING DATE: 1998-10-20
7	PRIOR APPLICATION NUMBER: 60/105002
7	PRIOR FILING DATE: 1998-10-20
7	PRIOR APPLICATION NUMBER: 60/105104
7	PRIOR FILING DATE: 1998-10-21
7	PRIOR APPLICATION NUMBER: 60/105169
7	PRIOR FILING DATE: 1998-10-22
7	PRIOR APPLICATION NUMBER: 60/105266
7	PRIOR FILING DATE: 1998-10-22
7	PRIOR APPLICATION NUMBER: 60/105693
7	PRIOR FILING DATE: 1998-10-26
7	PRIOR APPLICATION NUMBER: 60/105694
7	PRIOR FILING DATE: 1998-10-26

; PRIOR APPLICATION NUMBER: 60/105807

Alignment Scores:
Pred. No.: 0.447 Length: 451
Score: 113.50 Matches: 90
Percent Similarity: 33.16% Conservative: 38
Best Local Similarity: 23.32% Mismatches: 124
Query Match: 1.41% Indels: 134
DB: 10 Gaps: 23

US-10-776-827-10 (1-4460) x US-09-946-374-142 (1-451)

QY 318 GTGGCATGCCCAACCCAAAGCTGGAAGGATTAATACAGGTGCAAGTTGTCATC 377
Db ValHisCysProGlnPro-----ValThrGluProGlnGlnCysCys 92
QY 378 CTTTGTGAGCTCAGGACCTGCTTGT---AAGCCGAGAGGGTTCTCTGGC----- 422
Db -----ProLysCysValGluProHisThrProSerGlyLeuArgAlaPro 107
QY 423 CCTAATCTAGCCAAAGC-----CATGGAGAG----- 449
Db 108 ProLysSerCysGlnHisAsnGlyThrMetTyrGlnHisGlyGluIlePheSerAlaHis 127
QY 450 -----AATCAGTGCCTCTT-----CAG 467
Db 128 GluLeuPheProSerArgLeuProAsnGlnCysValLeuCysSerCysThrGluGlyGln 147
QY 468 CTCTATCTCGGACACCATCTTGACCAACCAAGGAGCATCTACACACCATCTGCTG 527
Db 148 IleTyrCysGly-----LeuThrThrCysProGlu-----Pro 158
QY 528 GGGAGTCCTCTCGGGCTGCCACTCTTGTGTGATCATCACCTCTTCTTCTGTTGCCA 587
Db 159 GlyCysPro-----AlaProLeuProLeuProAsp 168
QY 588 TTGCTCTGGAGCCCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 647
Db 169 SerCysCysGlnAlaCysLysAspGlu-----AlaSerGlnGlnSerAspGlu 185
QY 648 GAAGAA-----GAAGATGAAGAAGACCTCTGTCATCTCTCTCAACCCCAAGCT 695
Db 186 AppSerValGlnSerLeuHisGlyValArgHisProGlnAspProCysSerSerAspAla 205
QY 696 TCTCCAGATGGAAGAAGACCATCATCTGCTGTTTAGTAGGAGGAGGAGGAGGAGGAG 755
Db 206 Gly-Arg---LysArgGlyProGlyThrProAlaProThrGlyLeu-SerAlaProLeu 224
QY 756 CCTTCTGGGGCTAAGCCTCTCTGACCAACACACAGACATTTTCCAGAACCCCTGAATA 815
Db 224 erPhele-----ProArgHisPheArg----- 231
QY 816 ATGCATATGTCCATGTCTCAGAGTAATCTACTCAACC-----AAGG 857
Db 232 -----ProLysGlyAlaGlySerThrThrValLysIleValLeuLysG 246
QY 858 AACAACTCAGACTAAGTGTCCAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 917
Db 246 lu-LysHisLysLysAlaCysValHisGlyLysThrTyrSer---HisGlyGluVal 264
QY 918 TTGGATCTCTTGGCAGCTATGTGTCACATA---GCAATGTCTCTTACTGACAGCCCA 974
Db 265 TrpHisProAlaPheArgAlaPheGlyProLeuProCysIleLeuCysThrCysGluAsp 284
QY 975 GGCATGCTCTCC-----ACCTGTCTTGGCATATCCCAACATCCCAAGCAACAA 1022
Db 285 GlyArgGlnAspCysGlnArgValThrCysProThrGluTyrPro---CysArgHisPro 303
QY 1023 GAA-----CAT 1028
Db 304 GluLysValAlaGlyLysCysCysLysIleCysProGluAspLysAlaAspProGlyHis 323
QY 1029 TTATCCATACATCTCAATATGTTCCCAAGTGTGTGACATGCACGTAACACACACACAC 1088

Db 324 SerGluIleSerSerThrArgCysProLysAlaProGlyArgValLeuValHisThrSer 343
QY 1089 ACAATTCAGTAGCAGGTAGTGGCAAGTATATTCGTGCATCAATGTCATTTGGCT 1148
Db 344 ValSerProSerProAspAsnLeuArgArgPheAlaLeu----- 356
QY 1149 ATGTACTTTGTGACAGGAAAGTACATTATCTACAGTCAACAAATGTCTCATGGGAAAGCC 1208
Db 357 -----GluHisGluAlaSerAspLeuValGluIleTyrLeuTrpLysLeu 371
QY 1209 TTGCAGATTTCAGACACATATATCAATTTCTTACCAGCAGAGG-----CCCATACA 1262
Db 372 ValLysAspGluGluThrGluAlaGlnArgGlyGluValProGlyProArgProHisSer 391
QY 1263 CCATCTATTCCA 1274
Db 392 GlnAsnLeuPro 395
RESULT 12
US-10-206-915-282
; Sequence 282 Application US/10206915
; Publication No. US20040029221A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P34301C513
; CURRENT APPLICATION NUMBER: US/10/206,915
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: 10/052586
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059266
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063120
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063121
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063486
; PRIOR FILING DATE: 1997-10-21
; PRIOR APPLICATION NUMBER: 60/063540
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063541
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063544
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 282
; LENGTH: 451
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-206-915-282
Alignment Scores:
Pred. No.: 0.447 Length: 451
Score: 113.50 Matches: 90
Percent Similarity: 33.16% Conservative: 38
Best Local Similarity: 23.32% Mismatches: 124

Query Match:	1.41%	Indels:	134
DB:	12	Gaps:	23
US-10-776-827-10 (1-4460) x US-10-206-915-282 (1-451)			
QY	318	GTGGCATGCCCAACCCAAAGCTGGAAGAGGAATAAAATTACAAGTGGTCAAGGTTGCATC	377
DB	79	ValHisCysProGlnPro-----ValThrGluProGlnGlnCysCys	92
QY	378	CTTTTGTAGCTCAGACCTGCTTGT---AAGCCGAGAGGGTTCTCTGGC-----	422
DB	93	-----ProLysCysValGluProHisThrProSerGlyLeuArgAlaPro	107
QY	423	CCTAATCTAGCAACACAC-----CATGGAGAG-----	449
DB	108	ProLysSerCysGlnHisAsnGlyThrMetTyrGlnHisGlyGluIlePheSerAlaHis	127
QY	450	-----AATCAGTGCCTCTTT-----CAG	467
DB	128	GluLeuPheProSerArgLeuProAsnGlnCysValLeuCysSerCysThrGluGln	147
QY	468	CTCTATCTGGGACACCATCTTGCACCAACACCAAGAGGCATCTCAACACCATCTGGCT	527
DB	148	IleTyrCysGly-----LeuThrThrCysProGlu-----Pro	158
QY	528	GGGAGTCTCTCTGGGCTGGCCACTCTTGGTGATCATCATCTCTCTCTCTCTCTCTCT	587
DB	159	GlyCysPro-----AlaProLeuProLeuProAsp	168
QY	588	TTGCTGTGGAGCCACCAGGCAAGAGGGGGCCAGCAGCAGAGAGAGAGAGAGAGAGAG	647
DB	169	SerCysCysGlnAlaCysIleAspGlu-----AlaSerGluGlnSerAspGluGlu	185
QY	648	GAAGAA-----GAAGGATGAAGAGACCTCTGGATCTCTGTCTCAACCAAGCT	695
DB	186	AspSerValGlnSerLeuHisGlyValArgHisProGlnAspProCysSerSerAspAla	205
QY	696	TCTCAGATGAGAGAGACCATCATCTGCTGTTTAGTTAGTCAGGAGAGAGAGAGAGT	755
DB	206	Gly-Arg-----LysArgGlyProGlyThrProAlaProThrGlyLeu-SerAlaProLeu	224
QY	756	CCTTCTGGGGCTAAGCCTCTCTGACACACACAGACATTTACAGGAACCCCTCAATA	815
DB	224	erPheIle-----ProArgHisPheArg-----	231
QY	816	ATGCACTATGTCATGTCCACAGAGTAATCACTCAACC-----AAGG	857
DB	232	-----ProLysGlyAlaGlySerThrThrValLysIleValLeuLysG	246
QY	858	AACAAACCTCAGACTAAGTGTCCAGTGGAGGAGTCCAGGAGCAGCAGTGGCAATTC	917
DB	246	Iu-LysHisLysLysAlaCysValHisGlyGlyLysThrTyrSer---HisGlyGluVal	264
QY	918	TTGGATACTGTCTTGGCAGCTATGTGTCATAA---GCAATGCTCTTACTCTGACAGCCA	974
DB	265	TrpHisProAlaPheArgAlaPheGlyProLeuProCysIleLeuCysThrCysGluAsp	284
QY	975	GGCATGCTCCC-----ACCTGTCTCTGGCATACCCCATGCAAGACACAAA	1022
DB	285	GlyArgGlnAspCysGlnArgValThrCysProThrGluTyrPro---CysArgHisPro	303
QY	1023	GAA-----CAT	1028
DB	304	GluLysValAlaGlyLysCysLysIleCysProGluAspLysAlaAspProGlyHis	323
QY	1029	TTATCCATACATCTCAATATGTTCCCAAGTGTGTGCACATGCAGGTAAACACACACAC	1088
DB	324	SerGluIleSerSerThrArgCysProLysAlaProGlyArgValLeuValHisThrSer	343
QY	1089	ACAAATTCAGTAGCAGTACGTGGGCAAGTATATTTCTCTCATCAAAATGGTCTATGGCT	1148
DB	344	ValSerProSerProAspAsnLeuArgArgPheAlaLeu-----	356

QY	1149	ATGTACTTGTGCACGGAGTACATATCTACAGTCACAAAATGTCTCATCGGAAAGCC	1208
DB	357	-----GluHisGluAlaSerAspLeuValGluIleTyrLeuTrpLysLeu	371
QY	1209	TTGCAGATTCACACACATATATCAATTTCTCTCAACCAAGGAGG-----CCCCATACA	1262
DB	372	ValLysAspGluThrGluAlaGlnArgGlyGluValProGlyProArgProHisSer	391
QY	1263	CCATCTATTCCA	1274
DB	392	GlnAsnLeuPro	395
RESULT 13			
US-10-199-670-282			
; Sequence 282, Application US/10199670			
; Publication No. US20040033560A1			
; GENERAL INFORMATION:			
; APPLICANT: Baker, Kevin P.			
; APPLICANT: Chen, Jian			
; APPLICANT: Desnoyers, Luc			
; APPLICANT: Goddard, Audrey			
; APPLICANT: Godowski, Paul J.			
; APPLICANT: Gurney, Austin L.			
; APPLICANT: Pan, James			
; APPLICANT: Smith, Victoria			
; APPLICANT: Watanabe, Colin K.			
; APPLICANT: Wood, William I.			
; APPLICANT: Zhang, Zemin			
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC			
; FILE REFERENCE: P3430R1C401			
; CURRENT APPLICATION NUMBER: US/10/199,670			
; PRIOR FILING DATE: 2002-07-19			
; PRIOR APPLICATION NUMBER: 10/052586			
; PRIOR FILING DATE: 2002-01-15			
; PRIOR APPLICATION NUMBER: 60/059263			
; PRIOR FILING DATE: 1997-09-18			
; PRIOR APPLICATION NUMBER: 60/059266			
; PRIOR FILING DATE: 1997-09-18			
; PRIOR APPLICATION NUMBER: 60/062250			
; PRIOR FILING DATE: 1997-10-17			
; PRIOR APPLICATION NUMBER: 60/063120			
; PRIOR FILING DATE: 1997-10-24			
; PRIOR APPLICATION NUMBER: 60/063121			
; PRIOR FILING DATE: 1997-10-24			
; PRIOR APPLICATION NUMBER: 60/063486			
; PRIOR FILING DATE: 1997-10-21			
; PRIOR APPLICATION NUMBER: 60/063540			
; PRIOR FILING DATE: 1997-10-28			
; PRIOR APPLICATION NUMBER: 60/063541			
; PRIOR FILING DATE: 1997-10-28			
; PRIOR APPLICATION NUMBER: 60/063544			
; PRIOR FILING DATE: 1997-10-28			
; PRIOR Application data removed - See File Wrapper or PALM.			
; NUMBER OF SEQ ID NOS: 612			
; SEQ ID NO 282			
; LENGTH: 451			
; TYPE: PRT			
; ORGANISM: Homo Sapien			
US-10-199-670-282			
Alignment Scores:			
Pred. No.: 0.447			
Score: 113.50			
Length: 451			
Matches: 90			
Percent Similarity: 33.16%			
Best Local Similarity: 23.32%			
Query Match: 1.41%			
DB: 12			
US-10-776-827-10 (1-4460) x US-10-199-670-282 (1-451)			
QY	318	GTGGCATGCCCAACCCAAAGCTGGAAGAGGAATAAAATTACAAGTGGTCAAGGTTGCATC	377


```

Db      206 Gly-Arg---LysArgGlyProGlyThrProAlaProThrGlyLeu-SerAlaProLeuS 224
QY      756 CTTTCTGGGGCTAAGCCTCTCTGACCACACACAGACATTTCCAGAAACCCCTGAAATA 815
Db      224 erPheIle-----ProArgHisPheArg-----231
QY      816 ATGCACATATGCCATGTCCACAGAGTAATACTCAACC-----AAGG 857
Db      232 -----ProLysGlyAlaGlySerThrValLysIleValLeuLysG 246
QY      858 AACAACTCAGACTAAGTCCAGTCCAGTCCAGGAGTCCAGGAGCACCAGTGGCAAAATTC 917
Db      246 Iu-LysHisLysLysAlaCysValHisGlyGlyLysThrTyrSer---HisGlyGluVal 264
QY      918 TTGGATACTGTCTTGGCAGCTATGTCCAATA---GCAATGCTCCTTACTGCAGACCCA 974
Db      265 TrpHisProAlaPheArgAlaPheGlyProLeuProCysIleLeuCysThrCysGluAsp 284
QY      975 GGCATGCCCTCCC-----ACCTGTCTCTGGCATACCCACATGCCAAGCACAAA 1022
Db      285 GlyArgGlnAspCysGlnArgValThrCysProThrGluTyrPro---CysArgHisPro 303
QY      1023 GAA-----CAT 1028
Db      304 GluLysValAlaGlyLysCysCysLysIleCysProGluAspLysAlaAspProGlyHis 323
QY      1029 TTATCCATACATCTCAATATGTTCCCAAGTGTGTGCACATGCACGTAACACACACAC 1088
Db      324 SerGluIleSerSerThrArgCysProLysAlaProGlyArgValLeuValHisThrSer 343
QY      1089 ACAATTCAGGTAGCAGGTACGTGGGCAAGTATATTTCTCTCATCAATGGTCATGGCT 1148
Db      344 ValSerProSerProAspAsnLeuArgArgPheAlaLeu-----356
QY      1149 ATGTACTTTTGCAGGGAAGTACATTATCTACAGTCACAAAATGTCTCATGGGAAAGCC 1208
Db      357 -----GluHisGluAlaSerAspLeuValGluIleTyrLeuTyrLysLeu 371
QY      1209 TTGCCAGATTTCAGACACATATATACAAATTCCTTAACCAAGGC-----CCCCATACA 1262
Db      372 ValLysAspGluGluThrGluAlaGlnArgGlyGluValProGlyProArgProHisSer 391
QY      1263 CCATCTATTCCA 1274
Db      392 GlnAsnLeuPro 395

```

Search completed: September 21, 2004, 16:53:47
Job time : 498 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model
Run on: September 21, 2004, 15:10:50 ; Search time 72.5 Seconds
(without alignments)
6351.775 Million cell updates/sec

Title: US-10-776-827-10
Perfect score: 8026
Sequence: 1 cggggcagcaaccaggagat.....aattaaaggtagctgc 4460

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 389414 seqs, 51625971 residues
Total number of hits satisfying chosen parameters: 778828

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p.model -DEV=xl
-Q=/cgn2_1/USPTO_epool_p/US10776827/runat_21092004_133241_20811/app_query.fasta_1.4615
-DB=Issued Patents AA -QWMT=fastan -SUFFIX=n2p.ra -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEARSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10776827@cgn_1_119@runat_21092004_133241_20811 -NCPUS=6 -ICPU=3
-NO MMAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents AA:
1: /cgn2_6/ptodata/2/iaa/5A.COMB.pcp:*
2: /cgn2_6/ptodata/2/iaa/5B.COMB.pcp:*
3: /cgn2_6/ptodata/2/iaa/6A.COMB.pcp:*
4: /cgn2_6/ptodata/2/iaa/6B.COMB.pcp:*
5: /cgn2_6/ptodata/2/iaa/PCTUS.COMB.pcp:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pcp:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Match Length	Description
1	113.5	1.4	3 US-08-927-219-127
2	113	1.4	3 US-08-927-219-2
3	113	1.4	3 US-08-927-219-4
4	109	1.4	4 US-09-252-991A-31030
5	109	1.4	4 US-09-798-051-6
6	109	1.4	4 US-09-798-051-5
7	108.5	1.4	4 US-09-252-991A-30146
8	107	1.3	4 US-09-252-991A-23962
9	106.5	1.3	1 US-07-741-940-7
10	106.5	1.3	1 US-08-289-548A-7
11	106.5	1.3	1 US-08-452-654-7
12	106.5	1.3	1 US-08-449-731-7

13	106.5	1.3	2843	1	US-07-741-940-2	Sequence 2, Appli
14	106.5	1.3	2843	1	US-08-289-548A-2	Sequence 2, Appli
15	106.5	1.3	2843	1	US-08-452-654-2	Sequence 2, Appli
16	106.5	1.3	2843	1	US-08-452-655B-2	Sequence 2, Appli
17	106.5	1.3	2843	1	US-08-452-655B-7	Sequence 7, Appli
18	106.5	1.3	2843	2	US-08-370-235A-2	Sequence 2, Appli
19	106.5	1.3	2843	3	US-08-450-582-2	Sequence 2, Appli
20	106.5	1.3	2843	3	US-08-450-582-7	Sequence 2, Appli
21	106.5	1.3	2843	4	US-08-449-731-2	Sequence 7, Appli
22	106.5	1.3	2973	2	US-08-821-355A-7	Sequence 7, Appli
23	106.5	1.3	2973	2	US-09-003-687A-7	Sequence 7, Appli
24	106.5	1.3	2973	3	US-09-136-605-7	Sequence 7, Appli
25	105	1.3	1086	4	US-09-252-991A-25051	Sequence 25051, A
26	104.5	1.3	503	4	US-09-252-991A-32776	Sequence 32776, A
27	104	1.3	503	4	US-09-252-991A-18052	Sequence 18052, A
28	104	1.3	1184	4	US-09-266-225D-18	Sequence 18, Appli
29	104	1.3	5405	3	US-08-718-388-9	Sequence 9, Appli
30	102	1.3	530	4	US-09-800-729-112	Sequence 112, App
31	101.5	1.3	1185	3	US-09-041-886-23	Sequence 23, Appli
32	101	1.3	569	4	US-09-252-991A-24513	Sequence 24513, A
33	101	1.3	1940	2	US-08-644-271-30	Sequence 30, Appli
34	101	1.3	1940	4	US-09-077-955-34	Sequence 34, Appli
35	100.5	1.3	1187	1	US-08-320-559-28	Sequence 28, Appli
36	100.5	1.3	1187	3	US-08-545-860D-28	Sequence 28, Appli
37	100.5	1.3	1187	5	PCT-US94-04496-28	Sequence 28, Appli
38	100.5	1.3	1210	1	US-08-320-559-26	Sequence 26, Appli
39	100.5	1.3	1210	3	US-08-545-860D-26	Sequence 26, Appli
40	100.5	1.3	1210	5	PCT-US94-04496-26	Sequence 26, Appli
41	100	1.2	1257	1	US-08-340-428B-49	Sequence 49, Appli
42	99	1.2	224	4	US-09-252-991A-23849	Sequence 23849, A
43	98.5	1.2	558	4	US-09-252-991A-23318	Sequence 23318, A
44	98	1.2	431	4	US-09-252-991A-20267	Sequence 20267, A
45	98	1.2	589	1	US-07-668-648-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-927-219-127
; Sequence 127, Application US/08927219
; Patent No. 6187533
; GENERAL INFORMATION:
; APPLICANT: Bell, Graeme I.
; APPLICANT: Yamagata, Kazuya
; APPLICANT: Oda, Naohisa
; APPLICANT: Kaisaki, Pamela J.
; APPLICANT: Furuta, Hiroto
; APPLICANT: Horikawa, Yukio
; APPLICANT: Menzel, Stephen
; TITLE OF INVENTION: MUTATIONS IN THE DIABETES SUSCEPTIBILITY
; TITLE OF INVENTION: GENES HEPATOCYTE NUCLEAR FACTOR (HNF) 1 ALPHA, HNF-1BETA
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/927,219
; FILING DATE: Concurrently Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/029,679
; FILING DATE: 30-OCT-1996
; PRIOR APPLICATION DATA:

Wed Sep 22 17:04:10 2004

```

APPLICATION NUMBER: US 60/028,056
FILING DATE: 02-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/025,719
FILING DATE: 10-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Wilson, Mark B.
REGISTRATION NUMBER: 37,259
REFERENCE/DOCKET NUMBER: ARCD:272
TELECOMMUNICATION INFORMATION:
TELEPHONE: 512/418-3000
TELEFAX: 512/474-7577
INFORMATION FOR SEQ ID NO: 127:
SEQUENCE CHARACTERISTICS:
LENGTH: 631 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-927-219-127

Alignment Scores:
Pred. No.: 0.0169 Length: 631
Score: 113.50 Matches: 103
Percent Similarity: 31.76% Conservative: 45
Best Local Similarity: 22.10% Mismatches: 179
Query Match: 1.41% Indels: 139
DB: 3 Gaps: 23

US-10-776-827-10 (1-4460) x US-08-927-219-127 (1-631)
QY 225 GAGATCCCTGACACGAAAGACAGAGTAACGTGGAGGAGGTCCCAAGCTGGCTG 284
DB 193 GluLeuPro---ThrIysLysGlyArgArgAsnArgPheLysTrpGlyProAlaSerGln 211
QY 285 GCCATTTCTGCTGTGTGTGAGGTGCTGTAGTGGCATGCCAAACCCAAAGCTGGAA 344
DB 212 GlnIleLeuPheGlnAlaTyrglu---ArgGlnLysAsnProSerLysGlu 227
QY 345 GAGGAATAAATTACAAAGTGGTCAAGGTTGCATCTCTTTGAGTTCAGGACCTGCTTGTAG 404
DB 228 GluArg--GluThrLeuValGluGluCys-----AsnArgAlaGluCysIle 242
QY 405 CCAGAGGGTTCTGCGCTTAATCTAGCAACGACCATGGAGAGATCAGTGCCTTCTT 464
DB 243 GlnArgGlyValSer---ProSerGlnAlaGlnGlyLeuGlySerAsnLeuValThrGlu 261
QY 465 CAGCTCTATCTGGGACACCATCTTGACCAACACCAAGAGGATCTACAACACCATCTG 524
DB 262 ValArgValTyraAsnTrpPheAlaAsnArgArgLysGluGluAlaPheArgHisLysLeu 281
QY 525 -----CCTGGAGTCTCTCTGGGCTGCTGGGCTGCCACTCTTGGT 557
DB 282 AlaMetAspThrTyrsSerGlyProProGlyProGlyProGlyProGlyProAla---LeuPro 300
QY 558 GATCATCAGACTCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 617
DB 301 AlaHisSerSerPro---GlyLeuProProAlaLeuSerProSerLysValHisGly 319
QY 618 -----CCAGCAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGATGA 662
DB 320 ValArgTyrglyGlnProAlaThrSerGluThrAlaGluValProSerSerSerGly--- 338
QY 663 AGAAGACCTCTGGATCTCTGCTCAACC-----CAAGCTTCCAGATGAGAGAGAG 713
DB 339 ---GlyProLeuValThrValSerThrProLeuHisGlnValSerProThrGlyLeuGlu 357
QY 714 ACCATCATGCTGCTTTAGTAGCAGCAGCAGAGGTGTTCTCTTCTGGGGCTAAGCC 773
DB 358 Pro---SerHisSerLeuLeuSerThrGluAlaLysLeuValSerAlaAlaGly----- 374
QY 774 TCCTTCTGACACACAGACATTTCCAGAACCCCTGAAATATGACTATGTCCTCATGTC 833
DB 375 -----GlyProLeuProValSerThrLeuThrAlaLeu 386

```

```

834 CACAGA-----GTAACACTACTCAACCAAGAAACAACCTCAGACT 872
DB 387 HisSerLeuGluGlnThrSerProGlyLeuAsnGlnGlnProGlnAsnLeuMetAla 406
QY 873 AAGTGTCCAGTGGAGGCGAGTCCAGGAGCA----- 905
DB 407 SerLeuPro---GlyValMetThrIleGlyProGlyGluProAlaSerLeuGlyProThr 425
QY 905 ----- 905
DB 426 PheThrAsnThrGlyAlaSerThrLeuValIleGlyLeuAlaSerThrGlnAlaGlnSer 445
QY 906 -----CGTGGACAATTCCTGGATATCTGCTTGGCAGCT- 938
DB 446 ValProValIleAsnSerMetGlySerSerLeuThrThrLeuGlnProValGlnPheSer 465
QY 939 -----ATGTGTCCAATAGCAATGCTCTCTTACT 965
DB 466 GlnProLeuHisProSerTyrglnGlnProLeuMetProValGlnSerHisValThr 485
QY 966 GCAGACCCAGGATGCTCCACCTGCTCTGGCATACCCACATGCA-----AAGCAC 1019
DB 486 GlnAsnProPheMetAlaThrMetAlaGlnLeuGlnSerProHisAlaLeuTyrsHis 505
QY 1020 AAGAACCAATTTATCC-----ATACATCTCAATATGTTCCCAAGTGTGCACATGAC 1073
DB 506 LysProGluValAlaGlnTyThrHisThrGlyLeuLeuProGlnThr-----MetLeu 523
QY 1074 GTAACACACACACACAAATTCAGTAGCAGGTAGTGGGCAAGTATATCTGCTCATC 1133
DB 524 IleThrAspThr----- 527
QY 1134 AAATGCTCATTTGGCTATGTTTGTGCGAGGAGTACATATATCTACAGTCACAAAATG 1193
DB 528 -----ThrAsnLeuSerAlaLeuAlaSerLeu 536
QY 1194 TCTCATGGAAAGCCTTGCAGATTCAGACATATATACAAATTCCTAACGACGAGGC 1253
DB 537 ThrProThrLysGlnValPheThrSerAspThr-----GluAlaSerSerGluSerGly 554
QY 1254 CCCCATACACCATCTATTCATTAACCACTCAGGTACAGATGCATCTTCTCTTCTTCT 1313
DB 555 LeuHisThrProAlaSerGlnAlaThrThrLeuHisValProSerGlnAspProAlaGly 574
QY 1314 -----AACTCTACATAAATCTTTACTGGAAGTACTCATATAATTGGACATTC 1361
DB 575 IleGlnHisLeuGlnProAlaHisArgLeuSerAlaSerProThrValSerSerSer 594
QY 1362 AGCAACCTGCTACAGTCC 1379
DB 595 LeuValLeuTyrglnSer 600

RESULT 2
US-08-927-219-2
; Sequence 2, Application US/08927219
; Patent No. 6187533
; GENERAL INFORMATION:
; APPLICANT: Bell, Graeme I.
; APPLICANT: Yamagata, Kazuya
; APPLICANT: Oda, Naohisa
; APPLICANT: Katsuki, Pamela J.
; APPLICANT: Furuta, Hiroko
; APPLICANT: Horikawa, Yukio
; APPLICANT: Menzel, Stephen
; TITLE OF INVENTION: MUTATIONS IN THE DIABETES SUSCEPTIBILITY
; TITLE OF INVENTION: GENES HEPATOCYTE NUCLEAR FACTOR (HNF) 1 ALPHA, HNF-1BETA
; TITLE OF INVENTION: AND HNF-4ALPHA
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston

```

STATE: Texas
 COUNTRY: USA
 ZIP: 77210
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/927,219
 FILING DATE: Concurrently Herewith
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 60/029,679
 FILING DATE: 30-OCT-1996
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 60/028,056
 FILING DATE: 02-OCT-1996
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 60/025,719
 FILING DATE: 10-SEP-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Wilson, Mark B.
 REGISTRATION NUMBER: 37,259
 REFERENCE/DOCKET NUMBER: ARCD:272
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 512/418-3000
 TELEFAX: 512/474-7577
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 630 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-927-219-2

Alignment Scores:
 Pred. No.: 0.019 Length: 630
 Score: 113.00 Matches: 103
 Percent Similarity: 31.6% Conservative: 44
 Best Local Similarity: 22.15% Mismatches: 180
 Query Match: 1.41% Indels: 138
 DB: 3 Gaps: 23
 US-10-776-827-10 (1-4460) x US-08-927-219-2 (1-630)

QY 225 GAGATCCCTGACACGCAAGCAAGTAAACGTGGAGGAGGCTCCCAAGCTGGCTG 284
 D 193 GluLeuPro---ThrLysLysGlyArgAsnArgPheLysTrpGlyProAlaSerGln 211
 QY 285 GCCATTGCTGCTGCTGTGGAGTGCTGTCAAGTGTGATCTTTGAGTCTGAGGACCTGTTGAAG 344
 D 212 GlnLeuPheGlnAlaTyrGlu-----ArgGlnLysAsnProSerLysGlu 227
 QY 345 GAGGAATAAATACAAAGTGTCAAGTTGTCATCTTTGAGTCTGAGGACCTGCTGTGAAG 404
 D 228 GluArg---GluThrLeuValGluGluCys-----AsnArgAlaGluCysIle 242
 QY 405 CCGAGAGGTCTCTGGCCCTTAATCTAGCCAAAGCACCATGAGAGAAATCACTGCTTCTT 464
 D 243 GlnArgGlyValSer---ProSerGlnAlaGlnGlyLeuGlySerAsnLeuValThrGlu 261
 QY 465 CAGCTCTATCTGGGACACCATCTTGACCAACACCAAGAGGCTCTACAAACACCATCTG 524
 D 262 ValArgValTyrAsnTrpPheAlaAsnArgArgLysGluAlaPheArgHisLysLeu 281
 QY 525 -----CCTGGAGTCTCTGCTGGGCTGCTGGGCTGCTGGCTGCTGGT 557
 D 282 AlaMetAspThrTyrSerGlyProProGlyProGlyProGlyProGlyProAla---LeuPro 300
 QY 558 GATCATCACATCTCTTCTATCTGTGTCATCTGTGTCATCTGTGTCATCTGTGTCATCTGTG 617
 D 301 AlaHisSerSerPro---GlyLeuProProProAlaLeuSerProSerLysValHisGly 319

QY 618 -----COAGCAGCCAG 665
 D 320 ValArgGlyGlnProAlaThrSerGluThrAlaGluValProSerSerSerGly----- 337
 QY 666 AGACCTCTCGATCTCTGCTCAACC-----CAAGCTTCTCCAGATGGAGAGAGACC 716
 D 338 GlyProLeuValThrValSerThrProLeuHisGlnValSerProThrGlyLeuGluPro 357
 QY 717 ATCACTGCTCTGTTAGTTAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 776
 D 358 ---SerHisSerLeuLeuSerThrGluAlaLysLeuValSerAlaAlaGly----- 373
 QY 777 TTCTGACCACACAGACATTTTCAGGAACCCCTGAATAATGCATCTATCTCCATGTCCAC 836
 D 374 -----GlyProLeuProProValSerThrLeuThrAlaLeuHis 386
 QY 837 AGA-----GTAACACTACTCAACAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 875
 D 387 SerLeuGluGlnThrSerProGlyLeuAsnGlnProGlnAsnLeuLeuMetAlaSer 406
 QY 876 TGTCCAGTGGAGGAGTCCAGGAGCA----- 905
 D 407 LeuPro---GlyValMetThrIleGlyProGlyGluProAlaSerLeuGlyProThrPhe 425
 QY 905 ----- 905
 D 426 ThrAsnThrGlyAlaSerThrLeuValIleGlyLeuAlaSerThrGlnAlaGlnSerVal 445
 QY 906 -----CGTGGACAATCTTGGATCTGCTGGCAGCT----- 938
 D 446 ProValIleAsnSerMetGlySerSerLeuThrThrLeuGlnProValGlnPheSerGln 465
 QY 939 -----ATGTGTCCAATAGCAATGCTCTTACTGCA 968
 D 466 ProLeuHisProSerTyrGlnGlnProLeuMetProProValGlnSerHisValThrGln 485
 QY 969 GACCCAGGACATGCTCCACCTGCTCTGGCATACCCACCATGCA-----AAGCACA 1022
 D 486 SerProPheMetAlaThrMetAlaGlnLeuGlnSerProHisAlaLeuTyrSerHisLys 505
 QY 1023 GAACATTTATCC-----ATACATCTCAATATGTTCCCAAGTGTGTCACATGACGTA 1076
 D 506 ProGluValAlaGlnTyrThrHisThrGlyLeuLeuProGlnThr-----MetLeuIle 523
 QY 1077 ACACACACACACACAAATTCAGGTAGCAGGTACGTGGGCAAGTATATTTCTGCTCATCAA 1136
 D 524 ThrAspThr----- 526
 QY 1137 TGGTCATTGGCTATGTACTTTTGTGAGGAGAGTACATATCTACAGTCACAAAATGTCT 1196
 D 527 -----ThrAsnLeuSerAlaLeuAlaSerLeuThr 536
 QY 1197 CATGGGAAGCCTTCCAGATTCAGACACATATATACAAATTTCTAACCAGCAAGGCC 1256
 D 537 ProThrLysGlnValPheThrSerAspThr-----GluAlaSerSerGluSerGlyLeu 554
 QY 1257 CATACACATCTATTTCATAAACCACTCAGGTTACAGATGCATGCTTCTCTATTCTCT 1313
 D 555 HisThrProAlaSerGlnAlaThrLeuHisValProSerGlnAspProAlaGlyIle 574
 QY 1314 -----AACTCTACACATAAACTTTTACTGGAGTACTCATATAATTGGACATTTCCAGC 1364
 D 575 GlnHisLeuGlnProAlaHisArgLeuSerAlaSerProThrValSerSerSerLeu 594
 QY 1365 AACCTGTACAGTCC 1379
 D 595 ValLeuTyrGlnSer 599

RESULT 3
 US-08-927-219-4
 ; Sequence 4, Application US/08927219
 ; Patent No. 6187533

QY	405	CCGAGAGGGTCTCTGGCCCTTAATCTACCAAGCACCATCGAGAGAATCAGTGCCCTCTCTT	464
DB	243	GlnArgGlyValSer---ProSerGlnAlaGlnGlyLeuGlySerAsnLeuValThrGlu	261
QY	446	CAGCTCTATCTGGGACACCATCTTGACAAACAACCAAGAAGGCATCTACAAACCATTGTG	524
DB	262	ValArgValTyrrAsnTrpPheAlaAsnArgLysGluGluAlaPheArgHisLysLeu	281
QY	525	-----CCTGGGAGTCCTCTGGCCCTGCCACTCTCTTGGT	557
DB	282	AlaMetAspThrTyrrSerGlyProProGlyProGlyProGlyProAla---	300
QY	558	GATCATCACACCTCTTCATCTGTGTCATCTGTCTGGAGCCCACCAGCAAGAGGGG	617
DB	301	AlaHisSerPro---GlyLeuProProAlaLeuSerProSerLysValHisGly	319
QY	618	-----CCAGCAGCCAGAGAGAAAAGAGAGAGAGAGAGAGAGATGAAGA	665
DB	320	ValArgGlyGlnProAlaThrSerGluThrAlaGluValProSerSerSerGly----	337
QY	666	AGACCTCTGGATCTCTGCTCAACC-----CAAGCTTCTCCAGATGGAGAGAGAC	716
DB	338	GlyProLeuValThrValSerThrProLeuHisGlnValSerProThrGlyLeuGluPro	357
QY	717	ATCATCTGCCTGTTTAGTTAGGCAGGAAGCAGAGGTGTTTTCTTCTGGGGGTAAAGCCTCC	776
DB	358	---SerHisSerLeuLeuSerThrGluAlaLysLeuValSerAlaAlaGly-----	373
QY	777	TTCGTGCCACACAGACGATTTTCAGGAACCCCTGAAATAATCCACTATGTCCTGCCAC	836
DB	374	-----GlyProLeuProProValSerThrLeuThrAlaLeuHis	386
QY	837	AGA-----GTAACACTCAACCAAGAACAAACCTCAGACTAAG	875
DB	387	SerLeuGluGlnThrSerProGlyLeuAsnGlnProGlnAsnLeuIleMetAlaSer	406
QY	876	TGTCCTCAGTGGAGGGCAGTCCAGGACCA-----	905
DB	407	LeuPro--GlyValMetThrIleGlyProGlyGluProAlaSerLeuGlyProThrPhe	425
QY	905	-----	905
DB	426	ThrAsnThrGlyAlaSerThrLeuValIleGlyLeuAlaSerThrGlnAlaGlnSerVal	445
QY	906	-----COTGGACAATCTTGATCTGCTTGGCAGCT-----	938
DB	446	ProValIleAsnSerMetGlySerSerLeuThrThrLeuGlnProValGlnPheSerGln	465
QY	939	-----ATGTGTCCAATAGCAATGTCTTACTGCA	968
DB	466	ProLeuHisProSertyrGlnGlnProLeuMetProProValGlnSerHisValThrGln	485
QY	969	GACCCAGGCATGCTCCACCTGCTCTGTGCATACCCCCACATCCA-----AAGCACAAA	1022
DB	486	SerProPheMetAlaThrMetAlaGlnLeuGlnSerProHisAlaLeutyrrSerHisLys	505
QY	1023	GAACATTTATCC-----ATACATCTCAATATGGTTCCTCCAAAGTGTGTCACATGCACGTA	1076
DB	506	ProGluValAlaGlnTyrrThrHisThrGlyLeuLeuProGlnThr-----MetLeuIle	523
QY	1077	ACACACACACACAAAAATTCAGGTACGAGGTACGTGGGCAAGTATATTTCTGCTCATCAAA	1136
DB	524	ThrAspThr-----	526
QY	1137	TGGTCAITGGCTATGTACTTTGCGGGAGTAGTACATTATCTACGTCACAAAATGCTCT	1196
DB	527	-----ThrAsnLeuSerAlaLeuAlaSerLeuThr	536
QY	1197	CATGGAAAGCCTTGCCAGATTCCAGACATATATACAATTTCTCTAACCAGCAAGGCC	1256
DB	537	ProThrLysGlnValPheThrSerAspThr-----GluAlaSerSerGluSerGlyLeu	554

QY 1257 CATACACCATCTATTCATTAACCACTCAGGTTACAGATGATGCTTTCTCTATTTCT--- 1313
 Db |||||
 555 HisThrProAlaSerGlnAlaThrThrLeuHisValProSerGlnAspProAlaGlylle 574
 QY 1314 -----AACTCTACACATAAATCTTTTACTGGAAGTACTCTAATTTGGACATTCAGC 1364
 Db |||||
 575 GlnHisLeuGlnProAlaHisArgLeuSerAlaSerProThrValSerSerSerLeu 594
 QY 1365 AACCTGCTACAGTCC 1379
 Db |||||
 595 ValLeuTyrglnSer 599

RESULT 4
 US-09-252-991A-31030
 ; Sequence 31030, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 31030
 ; LENGTH: 180
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-31030

Alignment Scores:
 Pred. No.: 0.0233 Length: 180
 Score: 109.00 Matches: 50
 Percent Similarity: 39.46% Conservative: 23
 Best Local Similarity: 27.03% Mismatches: 62
 Query Match: 1.36% Indels: 50
 DB: Gaps: 9

US-10-776-827-10 (1-4460) x US-09-252-991A-31030 (1-180)
 QY 206 CAGGAAGCAGTCTGGCCGAGATCCCTGCAACCGCAAGCAAGTAAGTGAAGG 265
 Db |||||
 29 GlnGlySerArgAlaGlySerGlnProLeuSerLysGlnAlaSer----- 44
 QY 266 AGGCTCCCAAGCTGGCTGGCCATTTTGTGCTGTGTGTGGAGGTGCTGTCAGTGGCATG 325
 Db |||||
 45 -----ArgSerAlaSerHisThrAspSerAlaSerThrGlySerValGly----- 59
 QY 326 CCCAAACCAAGCTGGAAGGAGGAATAATTACAAG---TGTCGAAGTTGCATCCTTTT 382
 Db |||||
 60 -----SerCysGlyArgSerValSerlleGlnAlaArgSerArgValSerSer--- 75
 QY 383 GAGCTCAGGACCTGCTGTGAAGCCGAGAGGGTTCTGCGCCCTAATCTAGCCACACCA 442
 Db |||||
 76 -----SerSerArgCysMetSerProIleSerArgSerArgArg 88
 QY 443 TGGAGAGAATCAGTGCCCTTTCTACCTCTATCTGGGACACCATCTTGACCAACACCAAG 502
 Db |||||
 89 TrpArgAsnCysSerGlyGlySerSerLeuSerTrpAla-----MetArgArg 104
 QY 503 AAGGATCTACACACCATCTGCTGGGAG----- 532
 Db |||||
 105 ArgAlaSerGlyValArgSerSerTrpAspArgValArgSerSerArgCysTrpAla 124
 QY 533 -----TCCCTCCGGCCCTGCCACTCTTGGTGATCATCATCTCTCT 574
 Db |||||
 125 ArgLeuArgArgArgSerAlaMetCysSerArgAlaTrpAlaSerSerProSerSer 144
 QY 575 TCATCTGTGTCATTGCTGTGGAGCCCAAGGAGGGCCAGGAGGGCCAGGACCAAGAGA 634

Db |||||
 145 SerArgAla---AlaSerGlyAla-----AlaArgGlyValLeuSerSerSer 161
 QY 635 AAAAGAAGAAGAAGAAGATGAAGAAGACCTCTCGATCTCTGCTCAACCCAAAGC 694
 Db |||||
 162 AlaArgAlaSerAlaArgSerArgLysArgPheSerGlyThr----- 176
 QY 695 TTCTCCAGATGGAGA 709
 Db |||||
 177 ---ThrArgTrpArg 180

RESULT 5
 US-09-798-051-6
 ; Sequence 6, Application US/09798051
 ; Patent No. 6632780
 ; GENERAL INFORMATION:
 ; APPLICANT: Zhang, Ke
 ; APPLICANT: Cam, Linh
 ; APPLICANT: Nakayama, Naoki
 ; TITLE OF INVENTION: Chordin-Like-2 Molecules and Uses Thereof
 ; FILE REFERENCE: 01-005
 ; CURRENT APPLICATION NUMBER: US/09/798,051
 ; CURRENT FILING DATE: 2001-03-05
 ; NUMBER OF SEQ ID NOS: 21
 ; SOFTWARE: Patent in ver. 2.0
 ; SEQ ID NO 6
 ; LENGTH: 408
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-798-051-6

Alignment Scores:
 Pred. No.: 0.0383 Length: 408
 Score: 109.00 Matches: 90
 Percent Similarity: 31.70% Conservative: 33
 Best Local Similarity: 23.20% Mismatches: 130
 Query Match: 1.36% Indels: 135
 DB: Gaps: 23

US-10-776-827-10 (1-4460) x US-09-798-051-6 (1-408)
 QY 318 GTGGCATGCCCAACCCAAAGCTGGAAGGAGTAATAATACAGTGGTCAAGTTGCATC 377
 Db |||||
 58 ValHisCysProGlnPro-----ValThrGluProGlnGlnCysCys 71
 QY 378 CTTTGTAGCTCAGGACCTGCTGTGT---AAGCGAGAGGGTTCTCTGGC----- 422
 Db |||||
 72 -----ProLysCysValGluProHisThrProSerGlyLeuArgAlaPro 86
 QY 423 CCTAATCTAGCCAAAGCAC-----CATGGAGAG----- 449
 Db |||||
 87 ProLysSerCysGlnHisAsnGlyThrMetTyrglnHisGlyGluIlePheSerAlaHis 106
 QY 450 -----AATCAGTGCCTCTTT-----CAG 467
 Db |||||
 107 GluLeuPheProSerArgLeuProAsnGlnCysValLeuCysSerCysThrGluGlyGln 126
 QY 468 CTCTATCTGGGACCATCTTTGACCAACACCAAGAGGCGCATCTACACACCATCTGCT 527
 Db |||||
 127 IleTyrglyGly-----LeuThrCysProGlu-----Pro 137
 QY 528 GGGAGTCTCTCTGGGCGCTGCCACTCTTGGTGATCATCACACTCTCTCTCTCTGTCG 587
 Db |||||
 138 GlyCysPro-----AlaProLeuProLeuProAsp 147
 QY 588 TTGCTGTGGAGCCCAAGAGAGGGCGCCAGCCAGAGAGAGAGAGAGAGAGAGAG 647
 Db |||||
 148 SerCysCysGlnAlaCysLysAspGlu-----AlaSerGluGlnSerAspGluGlu 164
 QY 648 GAAGAA-----GAAGGATGAGAGAGACCTCTGATCTCTCTCTCTCAACCCAGCT 695
 Db |||||
 165 AspSerValGlnSerLeuHisGlyValArgHisProGlnAspProCysSerSerAspAla 184


```

465 uArgAlaGlnGly-----AlaProAlaGlyThrLeuProGly---GlyArgAl 481
440 CCATCGGAGAGATCAGTGCCTTCTCAGCTCTATCTGGGACACCATCTTGACCAACACC 499
481 aHisGly-----GlnProAlaGlyArgArgAlaGlyPr 496
500 AAGAAAGGCATCTACAACACCATCTGCTGGAGTCTCTCTGGGCTTGGGCTTGGTGA 559
484 oArgArg-----GlnProAlaGlyArgArgAlaGlyPr 496
560 TCATCACACTCTCTTTCATCTCTTGGCTGCTGGAGCCACCCAGGCAAGAGGGGCC 619
496 oGluHis-----ProArgArgAlaGlyGly 505
620 AGCAGCCAGAGAGAAAGAGAGAGAGAGAGAGAGATGAAGACCTCTTGGATCT 679
505 yAlaAlaArgGlnArg-----ArgArgGlnValAspPr 517
680 CTGCTCAACCCCAAGCTTCTCCAGATGGAGAGACCATCACTGCCTTTAGTTAGGCA 739
517 oAlaAlaProAlaGlnArgProAlaArgArgAlaGlyThrProAlaAlaGlyArgArg 537
740 GGAACAGAGGTGTTTCTCTTCTGGGGCTAAGCCCTCTTCACACACACACACATTTTC 799
537 gGln-----ProAspProAspArgPr 544
800 AGGACCCCTGAATAATGCACATATGTCCATGTC-----CACAGAGTAACACTCTCA 850
544 oGlyArgProProAlaArgTyrArgLeuProAlaAlaGlyArgGlyAlaValProTrpG 564
851 ACCAAGGAACAAACCTCCAGACTTAAGTGTCCAGTGGAGGGCAGTCCCAGGGACCACTGG 910
564 nProAlaArgGlnProGln-----ProGlyGluArgArgAlaGlyArgArgGly 580
911 A 911
580 Y 580

RESULT 8
US-09-252-991A-23962
; Sequence 23962, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23962
; LENGTH: 281
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23962

Alignment Scores:
Pred. No.: 0.0494 Length: 281
Score: 107.00 Matches: 37
Percent Similarity: 40.00% Conservative: 23
Best Local Similarity: 24.67% Mismatches: 41
Query Match: 1.33% Indels: 49
DB: 4 Gaps: 6

US-10-776-827-10 (1-4460) x US-09-252-991A-23962 (1-281)

QY 202 CTTCCAGGAAGAGTTCTGGCCGAGATCCCTTCACACGCAAGCAAGTAACGTGG 261

```


QY 3401 ----TCCATGTTTATGCTCTCTGATTCCTCTGCTGCAACAGCAGCC---CC 3453
Db 2456 uGluSerAlaSerPheGluSerLeuSerProSerSer-----ArgProAlaSerPr 2473
QY 3454 AGCCCTTTATCTCTCCATTTTCATCTCTCCAGCTCTGTCCTGAACTGCCACTGGC 3513
Db 2473 oThrArgSerGlnAlaGlnThrProValLeuSerProSerLeuProAspMetSerLeuSe 2493
QY 3514 AACCCATGGGACC 3526
Db 2493 rThrHisSerSer 2497

RESULT 10

US-08-289-548A-7
; Sequence 7, Application US/08289548A
; Patent No. 5648212

GENERAL INFORMATION:

APPLICANT: ALBERTSEN, HANS
APPLICANT: ANAND, RAKESH
APPLICANT: CARLSON, MARY
APPLICANT: GRODEN, JOANNA
APPLICANT: HEDGE, PHILIP J.
APPLICANT: JOSLYN, GEOFF
APPLICANT: KINZLER, KENNETH
APPLICANT: MARKHAM, ALEXANDER F.
APPLICANT: NAKAMURA, YUSUKE
APPLICANT: THLIVERIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
NUMBER OF SEQUENCES: 102
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Allegretti, LTD
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/289,548A
FILING DATE: 12-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.46943
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 2842 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

ORGANISM: Homo sapiens

IMMEDIATE SOURCE:

CLONE: APC

US-08-289-548A-7

Alignment Scores:

Pred. No.: 0.228

Score: 106.50

Percent Similarity: 35.41%

Best Local Similarity: 24.59%

Query Match: 1.33%

DB: 1

Length: 2842

Matches: 75

Conservative: 33

Mismatches: 105

Indels: 92

Gaps: 14

US-10-776-827-10 (1-4460) x US-08-289-548A-7 (1-2842)

QY 2797 TCACATAACAGAAATAACTTGGCCATCTGCTGCACCAAAACCCAGGATGTGGAAGACATCT 2856
Db 2224 SerArgGlyArgThrMetIleHisIleProGlyValArgAsnSerSerSerThrSer 2243
QY 2857 CCCCAACTGCCACTGTCTCACAGACAAAGTGGCTTCTGTCCTCCACCTCTCTAGTCC 2916
Db 2244 ProValSerLysGlyProPro-----LeuLysThrProAlaSerLysSer 2259
QY 2917 CCTAGATGGATGGCTGGGAGAGGTGGAGCTGACAGCTGACAGCTAGTGTGCATAT 2976
Db 2260 ProSerGluGlyGlnThrAlaThrThrSerProArgGlyAlaLysProSerValLysSer 2279
QY 2977 GATCTA-----GGAGGGCGGATC----- 2994
Db 2280 GluLeuSerProValAlaArgGlnThrSerGlnIleGlyGlySerSerLysAlaProSer 2299
QY 2995 ---ACGGGATCCGGGACCATACAAAGTACATGGTTTCCATGGCAACTGCTTGTCTGTTT 3051
Db 2300 ArgSerGlySerArgAspSerThrProSerArgProAla----- 2312
QY 3052 GAATTAAGACAGCAGCTCAGTTCTCATTTGCCATGCAAGCCCT-----CTATCT 3099
Db 2313 -----GlnGlnProLeuSerArgProIleGlnSerProGlyArgAsnSerIleSer 2329
QY 3100 CCAGGCACAATGCTCCCTGCTGCTCTTAATCAATGAGCTTCTC-----TCA 3147
Db 2330 ProGlyArgAsnGlyIleSerProAsnLysLeuSerGlnLeuProArgThrSerSer 2349
QY 3148 CCCAGGGATGAACACCCAGAAACTCCTCT----- 3180
Db 2350 ProSerThrAlaSerThrLysSerSerGlySerGlyLysMetSerTyrThrSerProGly 2369
QY 3181 ---CAGTCACTTCCACAGCCGATGACTCAGAGAGCCAAACCCAGAAATGGGCGCTCTCT 3237
Db 2370 ArgGlnMetSerGlnGlnAsnLeuThrLysGlnThrGlyLeuSerLysAsnAlaSerSer 2389
QY 3238 TTCCCATCACAGACTCCCTCGACAACTTCTCTGGCGTAACCTAGAGAGTCCCAAGTGA 3297
Db 2390 IlePro-----ArgSerGluSerAla 2396
QY 3298 GGATAGGCCCTAAACGTTTTG-----TTAAATAACAGGTGCAT----- 3336
Db 2397 SerLysGlyLeuAsnGlnMetAsnGlyAsnGlyAlaAsnLysLysValGluLeuSer 2416
QY 3337 -----GAAAGGACCTAAGCCATTGTTG 3360
Db 2417 ArgMetSerSerThrLysSerSerGlySerGluSerAspArgSerGluArgProValLeu 2436
QY 3361 ATA-----TCCACTCTCTCTTCTTCCACTTCTCTCTATCTTTTC----- 3400
Db 2437 ValArgGlnSerThrPheIleLysGluAlaProSerProThrLeuArgArgLysLeuGl 2456
QY 3401 ----TCCATGTTTATGCTCTCTGATTCCTCTTCTGCTCCCTGACACAGACGCC---CC 3453
Db 2456 uGluSerAlaSerPheGluSerLeuSerProSerSer-----ArgProAlaSerPr 2473
QY 3454 AGCCCTTTATCTCTCTCCATTTTCACCTCTCCAGCTCTGTCCTCCAGCTCTGTCCTGAACTGCCACTGGC 3513
Db 2473 oThrArgSerGlnAlaGlnThrProValLeuSerProSerLeuProAspMetSerLeuSe 2493
QY 3514 AACCCATGGGACC 3526
Db 2493 rThrHisSerSer 2497

RESULT 11

US-08-452-654-7
; Sequence 7, Application US/08452654
; Patent No. 5691454
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS

us-10-776-827-10.n2p.ra1

Qy	2977	GATCTA-----		-----GAGGAGCGGAGC	
Db	2280	GluLeuSerProValAlaArgGlnThrSerGlnIleGlyGlySerSerLysAlaProSer			2299
Qy	2995	---ACCGGATCCGGACCATCAAGATAACATGGTTTCATGGCAACTGCTTCTCGTTT			3051
Db	2300	ArgSerGlySerArgAspSerThrProSerArgProAla-			2312
Qy	3052	GAATTAAGACAGCAGTCGATGTTTCATTCGATGCATGACAAGGCCT-			3099
Db	2313	-----GlnProLeuSerArgProIleGlnSerProGlyArgAsnSerIleSer			2329
Qy	3100	CCAGGCACATGTCCTCGTGTCTCTAATCCAAATGGACTGTGTC-			3147
Db	2330	ProGlyArgAsnGlyIleSerProProAsnLysLeuSerGlnLeuProArgThrSerSer			2349
Qy	3148	CCCAGGGATGAACACCCAGAACTCATCTCT-			3180
					2369

Qy	3181	---	CAGTCACTTCCACAGCGCATGACTCAGAAGAGGCCAAACCAGAAATGGGGGCTCTCTCT	3237
Db	2370	ArgGlnMetSerGlnGlnAsnLeuThrLysGlnThrGlyLeuSerLysAsnAlaSerSer	2389	
Qy	3238	TTCCCATTCACAGACTCCCTCGACACCTTCTCGCGTAACTATAGAGGAGTCCACAGTGA	3297	
Db	2390	IlePro-----ArgSerGluSerAla	2396	

[illegible][illegible]

RESULT 12
US-08-449-731-7
; Sequence 7, Application US/08449731
; Patent No. 6413727
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; DAYVICH

CARLSON, MARY
GRODEN, JOANNA

HEDGE, PHILIP J.
JOSLYN, GEOFF
KINZLER, KENNETH
MARKHAM, ALEXANDER F.
NAYAMURA, YUSUKE
THLIVERIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
GENE IN COLORECTAL CANCER IN HUMANS
NUMBER OF SEQUENCES: 102
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Allegretti, LTD
CENTRE, 1201, G Street NW

NAKAMURA, ISOBORI
 THLIVERIS, ANDREW
 TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
 GENE IN COLORECTAL CANCER IN HUMANS
 NUMBER OF SEQUENCES: 102
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Banner & Allegretti, LTD
 1401 G Street NW
 SUITE 200
 WASHINGTON, DC 20004-4400

Db 2330 ProGlyArgAsnGlyIleSerProAsnLysLeuSerGlnLeuProArgThrSerSer 2349
 QY 3148 CCCAGGATGAACACACCCAGAACTCACTTCT 3180
 Db 2350 ProSerThrAlaSerThrLysSerSerGlySerGlyLysMetSerTyrThrSerProGly 2369
 QY 3181 --CAGTCACCTCCACAGCCGATGACTCAGAAGAGCCAAACCCAGATGGGGCTCTCTT 3237
 Db 2370 ArgGlnMetSerGlnGlnAsnLeuThrLysGlnThrGlyLeuSerLysAsnAlaSerSer 2389
 QY 3238 TTCCCATCAGAGACTCCCTCGACACCTTTCTCTGGCGTAACCTAGAGAGCTCCACAGTGCA 3297
 Db 2390 IlePro 3336
 QY 3298 GGATAGGCCCTAAACGTTTTC 3336
 Db 2397 SerLysGlyLeuAsnGlnMetAsnAsnGlyAsnGlyAlaAsnLysLysValGluLeuSer 2416
 QY 3337 -----GAAAGGAGCCTAAGGCCATTGTTG 3360
 Db 2417 ArgMetSerSerThrLysSerSerGlySerGluSerAspArgSerGluArgProValLeu 2436
 QY 3361 ATA-----TCCACTCTCTTCTTCCACTTCCCTCTCTCATCTTTTC----- 3400
 Db 2437 ValArgGlnSerThrPheIleLysGluAlaProSerProThr-LeuArgArgLysLeuGl 2456
 QY 3401 -----TCCATGTTTATGCTCTCTCTGATTCCTCTTCTGCTGCACACAGACCGCC---CC 3453
 Db 2456 uGluSerAlaSerPheGluSerLeuSerProSerSer-----ArgProAlaSerPr 2473
 QY 3454 AGCCCTTATTCCTGCCATTTTCACCTCTCTCCAGCTCTGCTGCTGAACTGCCACTGCG 3513
 Db 2473 oThrArgSerGlnAlaGlnThrProValLeuSerProSerLeuProAspMetSerLeuSe 2493
 QY 3514 AACCCATGGGACC 3526
 Db 2493 rThrHisSerSer 2497
 RESULT 13
 US-07-741-940-2
 ; Sequence 2, Application US/07741940
 ; Patent No. 5352775
 ; GENERAL INFORMATION:
 ; APPLICANT: ALBERTSEN, HANS
 ; APPLICANT: ANAND, RAKESH
 ; APPLICANT: CARLSON, MARY
 ; APPLICANT: GRODEN, JOANNA
 ; APPLICANT: HEDGE, PHILIP J.
 ; APPLICANT: JOSLYN, GEOFF
 ; APPLICANT: KINZLER, KENNETH
 ; APPLICANT: MARKHAM, ALEXANDER F.
 ; APPLICANT: NAKAMURA, YUSUKE
 ; APPLICANT: THLIVERTIS, ANDREW
 ; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
 ; NUMBER OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
 ; NUMBER OF SEQUENCES: 94
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Banner, Birch, McKie & Beckett
 ; STREET: 1001 G Street, NW
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20001-4598
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/741.940
 ; FILING DATE: 19920109
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:

NAME:	Kagan, Sarah A.
REGISTRATION NUMBER:	32,141
REFERENCE/DOCKET NUMBER:	1107.035574
TELECOMMUNICATION INFORMATION:	
TELEPHONE:	202-508-9100
TELEFAX:	202-508-9299
INFORMATION FOR SEQ ID NO:	2:
SEQUENCE CHARACTERISTICS:	
LENGTH:	2843 amino acids
TYPE:	AMINO ACID
TOPOLOGY:	linear
MOLECULE TYPE:	protein
US-07-741-940-2	
Alignment Scores:	
Pred. No.:	0.228
Score:	106.50
Percent Similarity:	35.41%
Best Local Similarity:	24.59%
Query Match:	1.33%
DB:	1
US-10-776-827-10 (1-4460) x US-07-741-940-2 (1-2843)	
QY	2797 TCACATAACAGATAACTTGCATCTGCTGCACCAAAACCCAGGAGTGTGAAGACATCT 2856
Db	2225 SerArgGlyThrMetIleHisIleProGlyValAlaArgAsnSerSerSer 2244
QY	2857 CCCCACAACTGCCATGCTCCACAGAGCAAGCTGCCCTTCTCTCCACCTTCAGTCC 2916
Db	2245 ProValSerLysLysGlyProPro-----LeuLysThrProAlaSerLysSer 2260
QY	2917 CCTAGAAATGGATGGCTGGGAGAGGTGGAGGCTGACAGCTGACAGCTAGTGTCCAGATAT 2976
Db	2261 ProSerGluGlyGlnThrAlaThrThrSerProArgGlyAlaLysProSerValLysSer 2280
QY	2977 GATCTA-----GGAGGCGGATC----- 2994
Db	2281 GluLeuSerProValAlaArgGlnThrSerGlnIleGlyLysSerSerLysAlaProSer 2300
QY	2995 ---ACGGGATCCGGACACCTACAGTAACATGTTTCCATGGCAACTGCTTGCCTTT 3051
Db	2301 ArgSerGlySerArgAspSerThrProSerA-gProAla----- 2313
QY	3052 GAATTAAGACAGAGTCACTGTGTCAATGCCATGACAAAGGCTT-----CTATCT 3099
Db	2314 -----GlnGlnProLeuSerArgProIleGlnSerProGlyArgAsnSerIleSer 2330
QY	3100 CCAGGCACAATGTCCTGCTGTCTCTAATCCAAATGCTGCTC-----TCA 3147
Db	2331 ProGlyArgAsnGlyIleSerProProAsnLysLeuSerGlnLeuProArgThrSerSer 2350
QY	3148 CCCAGGATGAACACCCAGAACTCACTTCT----- 3180
Db	2351 ProSerThrAlaSerThrLysSerSerGlySerGlyLysMetSerTyrThrSerProGly 2370
QY	3181 ---CAGTCACTTCCACAGCCGATGACTCAGAGAGCAACCCAGAAATGGGCTCTCTT 2373
Db	2371 ArgGlnMetSerGlnGlnAsnLeuThrLysGlnThrGlyLeuSerLysAsnAlaSerSer 2390
QY	3238 TTCCCATCAGACTCCCTGACAACTTCTCTGGCGTAACCTAGAGGAGTCCAGTGC 3297
Db	2391 IlePro-----ArgSerGluSerAla 2397
QY	3298 GGATAGGCCCTAAACGTTTTC-----TTAATAACAGGTCAT----- 3336
Db	2398 SerLysGlyLeuAsnGlnMetAsnAsnGlyAsnGlyAlaAsnLysLysValGluLeuSer 2417
QY	3337 -----GAAAGGACCTTAAGCCATTTGTTG 3360
Db	2418 ArgMetSerSerThrLysSerSerGlySerGluSerAspArgSerGluArgProValLeu 2437


```

QY 2797 TCACATAACAGAACTTCCCATCTGCTGCACCAACCCAGGATGTGAAGACATCT 2856
Db 2225 SerArgGlyArgThrWetIleHisIleProGlyValArgAsnSerSerSerThrSer 2244
QY 2857 CCCCACAACTGCCACTGCTCACCAGGACAAAGTCCCTCTCTCCCACTCTCAGTCC 2916
Db 2245 ProValSerLysLysGlyProPro-----LeuLysThrProAlaSerLysSer 2260
QY 2917 CCTAGAAATGGATGGCTGGGAGAGGTGGAGGTGACAGCTGACAGCTGTCAGATAT 2976
Db 2261 ProSerGluGlyGlnThrAlaThrThrSerProArgGlyAlaLysProSerValLysSer 2280
QY 2977 GATCTA-----GGAGGGCGGATC----- 2994
Db 2281 GluLeuSerProValAlaArgGlnThrSerGlnIleGlySerSerLysAlaProSer 2300
QY 2995 ---ACCGGATCCGGACCACTAACATGTTTCCATGGCAACTGCTGCTCGTTT 3051
Db 2301 ArgSerGlySerArgAspSerThrProSerArgProAla----- 2313
QY 3052 GAATTAAGACAGCAGTGTGCTGATGCTGATGCTGATGCTGATGCTGATGCTGAT 3099
Db 2314 -----GlnGlnProLeuSerArgProIleGlnSerProGlyArgAsnSerIleSer 2330
QY 3100 CCAGGCACAAATGCTCCTGCTGCTCCTCAATCCATGCACTGCTC-----TCA 3147
Db 2331 ProGlyArgAsnGlyLeuSerProProAsnLysLeuSerGlnLeuProArgThrSerSer 2350
QY 3148 CCCCAGGGATGAACACCCAGAACTCACTTCT----- 3180
Db 2351 ProSerThrAlaSerThrLysSerSerGlySerGlyLysMetSerTyrThrSerProGly 2370
QY 3181 ---CAGTCACTTCCACAGCCGATGACTAGAGAGCCAAACCCAGAAATGGGCGCTCTCT 3237
Db 2371 ArgGlnMetSerGlnGlnAsnLeuThrLysGlnThrGlyLeuSerLysAsnAlaSerSer 2390
QY 3238 TTCCCATCACAGACTCCCTCGACACCTTCTCGCGTAACAGAGAGTCCAGTGA 3297
Db 2391 IlePro-----ArgSerGluSerAla 2397
QY 3298 GGATAGGCGCTAAACGTTTGT-----TTAAATAAACAGGTGCAT----- 3336
Db 2398 SerLysGlyLeuAsnGlnMetAsnAsnGlyAsnGlyAlaAsnLysLysValGluLeuSer 2417
QY 3337 -----GAAAGAGCGCTAAGGCCATGTTG 3360
Db 2418 ArgMetSerSerThrLysSerSerGlySerGluSerArgSerGluArgProValLeu 2437
QY 3361 ATA-----TCCACTCTCTCTTCCACTTCTCTCTCTCTCTCTCTCTCTCTCTCT 3400
Db 2438 ValArgGlnSerThrPheIleLysGluAlaProSerProThr-LeuArgLysLeuGln 2457
QY 3401 ---TCCAGTTTATGCTTCTCTGATTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3453
Db 2457 uGluSerAlaSerPheGluSerLeuSerProSerSer-----ArgProAlaSerPr 2474
QY 3454 AGCCCTTATTCCTCTCACTTTTCACTCTTCCAGCCTCTGTCCTGAACTGCCACTGGC 3513
Db 2474 oThrArgSerGlnAlaGlnThrProValLeuSerProSerLeuProAspMetSerLeuSe 2494
QY 3514 AACCCATGGGACC 3526
Db 2494 rThrHisSerSer 2498

```

RESULT 15

```

US-08-452-654-2
; Sequence 2, Application US/08452654
; Patent No. 5691454
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY

```

```

; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER P.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie & Beckett
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452,654
; FILING DATE: 25-MAY-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/741,940
; FILING DATE: 08-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.035574
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2843 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-452-654-2

```

Alignment Scores:

```

Pred. No.: 0.228 Length: 2843
Score: 106.50 Matches: 75
Percent Similarity: 35.41% Conservative: 33
Best Local Similarity: 24.59% Mismatches: 105
Query Match: 1.33% Indels: 92
DB: 1 Gaps: 14

```

US-10-776-827-10 (1-4460) x US-08-452-654-2 (1-2843)

```

QY 2797 TCACATAACAGAACTTCCCATCTGCTGCACCAACCCAGGATGTGAAGACATCT 2856
Db 2225 SerArgGlyArgThrWetIleHisIleProGlyValArgAsnSerSerSerThrSer 2244
QY 2857 CCCCACAACTGCCACTGCTCACCAGGACAAAGTCCCTCTCTCCCACTCTCAGTCC 2916
Db 2245 ProValSerLysLysGlyProPro-----LeuLysThrProAlaSerLysSer 2260
QY 2917 CCCTAGAATGGATGGCTGGGAGAGGTGGAGGTGACAGCTGACAGCTAGTGTGATAT 2976
Db 2261 ProSerGluGlyGlnThrAlaThrThrSerProArgGlyAlaLysProSerValLysSer 2280
QY 2977 GATCTA-----GGAGGGCGGATC----- 2994
Db 2281 GluLeuSerProValAlaArgGlnThrSerGlnIleGlySerSerLysAlaProSer 2300
QY 2995 ---ACCGGATCCGGACCACTAACATGTTTCCATGGCAACTGCTGCTCGTTT 3051
Db 2301 ArgSerGlySerArgAspSerThrProSerArgProAla----- 2313

```

Search completed: September 21, 2004, 15:48:20
Job time : 196.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: September 21, 2004, 15:06:59 ; Search time 410 Seconds
(without alignments)
6864.445 Million cell updates/sec

Title: US-10-776-827-10

Perfect score: 8026

Sequence: 1 cggggcagcaacaggagat.....aattaaggtatactagctc 4460

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xl
-Q=/cgn2_1/USPTO_spoop/US10776827/runat_21092004_133240_20778/app_query.fasta_1.4615
-DB=SPTRMBL_25 -QMT=fastan -SUFFIX=n2p.rsp -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10776827@cgn.1.1.728 @runat_21092004_133240_20778 -NCPU=6 -ICPU=3
-NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	403	5.0	126	11	Q8CE67	Q8CE67 mus musculus

2	214.5	2.7	160	11	Q8CBL2	Q8CBL2 mus musculus
3	174	2.2	62	4	Q8WW72	Q8WW72 homo sapien
4	171.5	2.1	122	11	Q9D4T5	Q9D4T5 mus musculus
5	129.5	1.6	1203	4	Q7Z3U0	Q7Z3U0 homo sapien
6	127.5	1.6	1203	4	Q9HC56	Q9HC56 homo sapien
7	125.5	1.6	1084	4	Q8NJK7	Q8NJK7 homo sapien
8	118	1.5	468	11	Q8BIR7	Q8BIR7 mus musculus
9	118	1.5	4998	11	Q8CG65	Q8CG65 mus musculus
10	114.5	1.4	468	11	Q99LH4	Q99LH4 mus musculus
11	114.5	1.4	1381	11	Q8CJ95	Q8CJ95 mus musculus
12	114.5	1.4	1400	11	Q8CGB6	Q8CGB6 mus musculus
13	114.5	1.4	1637	6	Q9XSV8	Q9XSV8 mus musculus
14	114.5	1.4	5146	6	Q8SPM4	Q8SPM4 mus musculus
15	113.5	1.4	451	4	Q7Z5J3	Q7Z5J3 homo sapien
16	112.5	1.4	473	11	Q8VDC2	Q8VDC2 mus musculus
17	111.5	1.4	1860	4	Q8IZC6	Q8IZC6 homo sapien
18	110	1.4	498	11	Q8CG65	Q8CG65 mus musculus
19	109	1.4	968	10	Q9SD31	Q9SD31 arabidopsis
20	108	1.3	614	13	Q80Z20	Q80Z20 brachydanio
21	108	1.3	1396	5	Q8T0T0	Q8T0T0 drosophila
22	107.5	1.3	786	10	Q9FMJ2	Q9FMJ2 arabidopsis
23	107.5	1.3	1210	11	Q811B5	Q811B5 rattus norv
24	107	1.3	841	11	Q9WVL8	Q9WVL8 mus musculus
25	107	1.3	971	4	Q96KW2	Q96KW2 homo sapien
26	107	1.3	981	11	Q8CHP6	Q8CHP6 mus musculus
27	106.5	1.3	964	4	Q8NFK1	Q8NFK1 homo sapien
28	106.5	1.3	983	4	Q8NDX5	Q8NDX5 homo sapien
29	106.5	1.3	983	4	Q8NFT7	Q8NFT7 homo sapien
30	106.5	1.3	1002	11	Q8BP69	Q8BP69 mus musculus
31	106.5	1.3	1055	11	Q8CHG2	Q8CHG2 mus musculus
32	106	1.3	382	16	Q8XVU9	Q8XVU9 mus musculus
33	106	1.3	655	11	Q8R122	Q8R122 mus musculus
34	106	1.3	981	11	Q8CBC0	Q8CBC0 mus musculus
35	106	1.3	981	11	Q8C881	Q8C881 mus musculus
36	106	1.3	989	11	Q9MLU9	Q9MLU9 mus musculus
37	105.5	1.3	1158	5	Q97169	Q97169 drosophila
38	105	1.3	735	4	Q95784	Q95784 homo sapien
39	104.5	1.3	1222	11	Q810H4	Q810H4 rattus norv
40	104	1.3	513	11	Q8R1U3	Q8R1U3 mus musculus
41	104	1.3	692	11	Q8BH09	Q8BH09 mus musculus
42	104	1.3	735	11	Q8K4N4	Q8K4N4 mus musculus
43	104	1.3	1419	4	Q9NFG0	Q9NFG0 homo sapien
44	103.5	1.3	330	4	Q14882	Q14882 homo sapien
45	103.5	1.3	1243	12	Q8JKK6	Q8JKK6 heliothis z

ALIGNMENTS

RESULT 1

Q8CE67
ID Q8CE67 PRELIMINARY; PRT; 126 AA.
AC Q8CE67;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Unclassifiable.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N. A.
RC STRAIN=C57BL/6J; TISSUE=Skin;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL: AK028922; BAC26196.1; -;
SQ SEQUENCE 126 AA; 14102 MW; CEB01EF3E6C50A44 CRC64;

Alignment Scores:

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 EN NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Human amygdala;
 RA Ansoerge W., Krieger S., Regiert T., Rittmueller C., Schwager B.,
 RL Mewes H.W., Weill B., Amid C., Osanger A., Fobo G., Han M., Wiemann S.,
 RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BX537422; CAD97664.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 1203 AA; 132214 MW; 25B905E7E6203F10 CRC64;

Alignment Scores:
 Pred. No.: 0.000931 Length: 1203
 Score: 129.50 Matches: 82
 Percent Similarity: 38.23% Conservative: 56
 Best Local Similarity: 22.71% Mismatches: 126
 Query Match: 1.61% Indels: 97
 DB: 4 Gaps: 18

US-10-776-827-10 (1-4460) x Q723U0 (1-1203)

QY	496	CACCAAGAGGCATCTACAACACCATCTGCCTGGAGTCCTCTGGCGCTGCCACACTCTTG	555
Db	807	TyrGlnAenGluAspTyrLeuThrIleMetIleAlaIleAlaGlyAlaMetValVal	826
QY	556	GTGATC-----ATCACACTCTCTTCATCTGTTGCCATTGCTGTC-----	594
Db	827	IleValValIlePheValThrValLeuValArgCysArgHisAlaSerArgPheLysAla	846
QY	595	-----TGGAGGCCACCAAGAGGCGCCAGCAGCCCA	627
Db	847	AlaGlnArgSerLysGlnGlyAlaGluTipMetSerProAsnGlnGluAsnLysGlnAsn	866
QY	628	GAGAGAAAAGAGAGAGACAGACAGACAGATGAGAGACTCT-----TGGATCTCT	681
Db	867	LysLysLysLysArgLysArgLysArgLysSerProLysSerSerLeuLeuAsnPheValThr	886
QY	682	GCTCAACCAAGCTTCTCCAGATGGA-----GAAGAGACCATCACTG	723
Db	887	IleGlu--GluSerLysProAspAspAlaValHisGluProIleAenGlyThrIleSerL	906
QY	724	CTGTGTTAGTTAGCAGGAAGCAGAGAGTGTTCCTTTC---TGGGCTTAAGCCTCCTCT	780
Db	906	euaAlaGluLeuGluGlnSerIleGlyArgPheAspTrpGlySerAlaPro-----	924
QY	781	GACCACACACAGACATTTCCAGAACCCCTGAATAATGCACTATGTCCATGTCCACAGAG	840
Db	925	--ProThrThrPheLysProAsnSerProAspLeuAlaLys-----HisTyrLysSerA	942
QY	841	TAATCTCAACCAAGGAACAAACCTCAGACTAAGTGTCCCACTGGAGGCGAGTCCCAGG	900
Db	942	laSerProGlnPro-AlaPheHisLeuLysProAspThrProVal-----SerValLys	959
QY	901	GACCAC-----TGGGCAATTCCTTGATACTGCTTCTGGCAGCTA	939
Db	960	LysHisHisValIleGlnGluLeuProLeuAspAsnThrPheValGlyGly-----	976
QY	940	TGTGTCCAATAGCAATGCTCTTACTTGCAGACCCAGGATGCTCC-----CACCT	990
Db	977	-----CysAspThrLeuSer-LysArgSerSerThrSerSerAspHisPh	991
QY	991	GTCTCTGGCATACCCACATGCAAGACCAAGAACATTT-----ATCCA	1035
Db	991	eSerAlaSer-----GluCysSerSerGlnGlyPheLysThrLysGlyProLeuHi	1009
QY	1036	TACATCTCATATATGGTTCCCAAGTGTGTGCATCATGCACTAACACACACACACAAATT	1095
Db	1009	sThrArgGlnSerGlnArgValThrPheHisLeuProAsp-----	1023
QY	1096	CAGGTAGCAGGTACGTGGGCAAGTATATCTGCTCATCAAAATGGTCAATGTGCTATGTA	1155

Db 1024 -----GlySerGlnGluSerCysSerAspSerGlyLeuGlyAspHisG1 1038
 QY 1156 TTGTGCGGGAAGTACATTATCTACAGTCCACAAAATGTCT----- 1196
 Db 1038 uProValGlySerGlyThrLeuLeuSerHis-ProLeuProLeuValGlnProGlnAspG 1058
 QY 1197 -----CATGGGAAAGCCTTGCAGATTCCAGACACATATATA-----CAATTCTCTA 1242
 Db 1058 luPheTyAspGlnAlaSerProAspLysArgThrGluAlaAspGlyAsnSerAspProA 1078
 QY 1243 ACCAGCAGGCCCCATACACATCTATCCATAAACCCACTCAGGTTACAGATCATGCT 1302
 Db 1078 nsSerAspGlyProLeuGlyProArg----- 1086
 QY 1303 TTCCTATTCTTAACTTACACATAAACCTTTTACTGGAAGTACTCATATAATGGACATTTCCA 1362
 Db 1087 --GlyLeuAlaGluAlaThrGluMetCysThrGlnGluCysLeuValLeuGlyHisSerA 1106
 QY 1363 GCAACTCTCTACAGTCCACCCACCTTGTGTCTTGTGATACAGACACACCAAG 1413
 Db 1106 spAsn-CysTrpMetPro---ProGlyLeuGlyProTyGlnHisProlys 1121

RESULT 6
 ID Q9HC56 PRELIMINARY; PRT; 1203 AA.
 AC Q9HC56;
 DT 01-MAR-2001 (TReMBLrel. 16, Created)
 DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
 DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
 DE Protocadherin-9.
 GN PCDH9.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID:9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kools P.F.J., Van Roy F.;
 RT Identification and characterization of a novel human protocadherin
 RT gene with high homology to a chromosome X-linked protocadherin.";
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF169692; AAF89689.2; -
 DR HSP; P15116; INCI.
 DR Genew; HGNC:8661; PCDH9.
 DR GO; GO:0016020; C-membrane; IEA.
 DR GO; GO:0005509; F:calcium ion binding; IEA.
 DR GO; GO:0005194; F:cell adhesion molecule activity; IEA.
 DR GO; GO:0007155; P:cell adhesion; IEA.
 DR GO; GO:0007156; P:homophilic cell adhesion; IEA.
 DR InterPro; IPR002126; Cadherin.
 DR Pfam; PF00028; cadherin; 6.
 DR PRINTS; PRO0205; CADHERIN.
 DR SMART; SM00112; CA; 7.
 DR PROSITE; PS00232; CADHERIN 1; 6.
 DR PROSITE; PS0268; CADHERIN 2; 7.
 DR Calcium; Calcium-binding; Cell adhesion; Glycoprotein.
 KW SEQUENCE 1203 AA; 132250 MW; CEEDD24FC1CF8BE2 CRC64;

Alignment Scores:
 Pred. NO.: 0.00151 Length: 1203
 Score: 127.50 Matches: 83
 Conserv: 55
 Percent Similarity: 38.12% Conservative: 125
 Best Local Similarity: 22.93% Mismatches: 99
 Query Match: 1.59% Indels: 18
 DB: 4 Gaps: 18

US-10-776-827-10 (1-4460) x Q9HC56 (1-1203)

QY 496 CACCAAGAGGCATCTACACACCATCTGCTGGAGTCTCTCTGGGCTGCCATCTTGT 555
 Db 807 TyrGlnAsnGluAspTyrlleuThrleMetlleAlaalleleAlaGlyAlaMetValVal 826

QY 556 GTGATC-----ATCACATCTCTTCTTCATCTGTGTCATCTGCTGC----- 594
 Db 827 lIleValIlePheValThrValLeuValArgCysArgHisAlaSerArgPheLysAla 846
 QY 595 -----TGAGCCCAACAGCAGCAGAGAGGGCGCAGACCCA 627
 Db 847 AlaGlnArgSerLysGlnGlyAlaGluTrpMetSerProAsnGlnGluAsnLysGlnAsn 866
 QY 628 GAGAAGAAAAGAAAGAAAGAAAGAGATGAGAAAGACCTC-----TGATCTCT 681
 Db 867 LysLysLysLysArgLysLysLysArgLysSerProLysSerSerLeuLeuAsnPheValThr 886
 QY 682 GCTCAACCCAGCTTCTCCAGAT-----GGAGAAGAGACCA 717
 Db 887 lIleGlu--GluSerLysProAspAspAlaValHisGluProIleAsnGlyThrIleSerL 906
 QY 718 TCATCTGCTGTTTAGTAGCAGCAGAGGTGTTCTCTTCTGGGGCTTAAGCCTCT 777
 Db 906 euProAlaGluLeuGluGlnSerIleGlyArgPheAsp--TipGlyProAlaPro- 924
 QY 778 TCTGACCACACACACATTTTCAGAACCCCTGAAATAATGCATATGTCATGTCACCA 837
 Db 925 -----ProThrThrPheLysProAsnSerProAspLeuAlaLys-----HisTyLysS 941
 QY 838 GAGTAATCTACCAACAGGAACAACCTCAGACTAAGTCCCACTGAGGGCAGTCCC 897
 Db 941 erAlaSerProGlnPro-AlaPheHisLeuLysProAspThrProVal-----SerVal 958
 QY 898 AGGAGCCAC-----GTGACAATTTCTGGATGACTGTCTTTGGCAG 936
 Db 959 LysLysHisHisValIleGlnGluLeuProLeuAspAsnThrPheValGlyGly----- 976
 QY 937 CTATGTGTCCATAGCAATGCTTACTTGACAGCCAGGATGCTCC-----CA 987
 Db 977 -----CysAspThrLeuSer-LysArgSerSerThrSerSerAspHi 990
 QY 988 CTGTCTGTGGCATCCCCACATGCAAGCACAAAGAACATTT-----AT 1032
 Db 990 sPheSerAlaSer-----GluCysSerSerGlnGlyGlyPheLysThrLysGlyProLe 1008
 QY 1033 CMTATATCTCATATGTTCCCAAGTGTGTCACATGCACATGACGTAACACACACACACAA 1092
 Db 1008 uHisThrArgGlnSerGlnArgArgValThrPheHisLeuProAsp----- 1023
 QY 1093 ATTCAGGTAGCAGGTACGTGGCAGCATATATCTGCTCATCAATGTCATTTGGCTATGT 1152
 Db 1024 -----GlySerGlnGluSerCysSerAspSerGlyLeuGlyAspHi 1037
 QY 1153 ACTTTGTGAGGAAGTACATTTATCTACATGTCACAAAATGTCT----- 1196
 Db 1037 sGluProValGlySerGlyThrLeuLeuSerHis-ProLeuProLeuValGlnProGlnA 1057
 QY 1197 -----CATGGGAAAGCCTTCCAGATTCCAGACACATATATA-----CAATTTC 1239
 Db 1057 spGluPheTyAspGlnAlaSerProAspLysArgThrGluAlaAspGlyAsnSerAspP 1077
 QY 1240 CTAACACAGCAGGCCCCCATACACATCTATTCATAAACACCTCAGGTTACAGATCAT 1299
 Db 1077 roAsnSerAspGlyProLeuGlyProArg----- 1086
 QY 1300 GCTTTCCTATTCTAACTCTACACATAAACTTTTACTGGAAGTACTCATATTTGGACATT 1359
 Db 1087 -----GlyLeuAlaGluAlaThrGluMetCysThrGlnGluCysLeuValLeuGlyHis 1105
 QY 1360 CCAGCACTCTCAGTCCGCCCTTGTGTGTCTTGTATACAGACACACACCAAG 1413
 Db 1105 erAspAsn-CysTrpMetPro---ProGlyLeuGlyProTyGlnHisProlys 1121

RESULT 7
 Q9N3K7 PRELIMINARY; PRT; 1084 AA.
 ID Q9N3K7
 AC Q9N3K7;
 DT 01-OCT-2002 (TReMBLrel. 22, Created)

[illegible]

[illegible]

Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.

RL EMBL: AJ491857; CAD42654.1; -
 DR GO: 0005576; C:extracellular; IEA.
 DR GO: 0005507; F:copper ion binding; IEA.
 DR GO: 0005489; F:electron transporter activity; IEA.
 DR GO: 0005179; F:hormone activity; IEA.
 DR GO: 0007155; P:cell adhesion; IEA.
 DR GO: 0006118; P:electron transport; IEA.
 DR InterPro: IPR000923; BlueCu_1.
 DR InterPro: IPR006207; Cys knot_C.
 DR InterPro: IPR000421; EGF-like.
 DR InterPro: IPR001545; Gly_hormoneB.
 DR InterPro: IPR002172; LDL_receptor_A.
 DR InterPro: IPR002919; TIL_Cysrich.
 DR InterPro: IPR000884; TSP1.
 DR InterPro: IPR008085; TSP_1.
 DR InterPro: IPR006552; VC_out.
 DR InterPro: IPR001007; VWF_C.
 DR Pfam: PF00754; F5_F8_type_C; 1.
 DR Pfam: PF00057; ldl_recept_a; 10.
 DR Pfam: PF01826; TIL; 10.
 DR Pfam: PF00090; tsp_1; 25.
 DR Pfam: PF00093; vwc; 1.
 DR Pfam: PF00094; vwd; 3.
 DR PRINTS: PR00261; LDLRECEPTOR.
 DR PRINTS: PR01705; TSP1REPEAT.
 DR SMART: SM00192; LDLa; 10.
 DR SMART: SM00209; TSP1; 25.
 DR SMART: SM00214; VMC; 4.
 DR SMART: SM00215; VMC_out; 9.
 DR SMART: SM00216; VWD; 2.
 DR PROSITE: PS00196; COPPER_BLUE; 1.
 DR PROSITE: PS01225; CTCK_2; 1.
 DR PROSITE: PS01186; EGF_2; 2.
 DR PROSITE: PS00022; FAS8C_3; 1.
 DR PROSITE: PS00261; GLYCO_HORMONE_BETA_1; 2.
 DR PROSITE: PS01209; LDLRA_1; 8.
 DR PROSITE: PS00068; LDLRA_2; 10.
 DR PROSITE: PS00092; TSP1; 25.
 DR PROSITE: PS01208; VWF_1; 1.
 DR PROSITE: PS01184; VWF_2; 2.
 SQ SEQUENCE 4998 AA; 535019 MW; DA2ABA8DA47DF225 CRC64;

 Alignment Scores:
 Pred. No.: 0.0249 Length: 4998
 Score: 118.00 Matches: 159
 Percent Similarity: 30.76% Conservative: 64
 Best Local Similarity: 21.93% Mismatches: 155
 Query Match: 1.47% Indels: 347
 DB: 11 Gaps: 45

 US-10-776-827-10 (1-4460) x Q8CG65 (1-4998)
 QY 1625 CAAGAGCTTTTGTCTTGTGTTGAGA--AAGGCATTGCTGAGTGTTC--GGG 1572
 Db 689 GlnGluGlyValLeuLeuGlnTrpArgAsnGlnThrLeuCysProValSerCysProGly 708
 QY 1571 GGTAAATGTTTCAGAGAGGCTTGGATCACATTAAACACCTCTCAGTGCAATGAGGAC 1512
 Db 709 GlyGlnValTyGlnGluCys-----AlaProValCysGly----- 720
 QY 1511 GGACTTGTCTATGAGTACAGAAAGAGCTTCAATGAATTTTGAACCATCACACTTTA 1452
 Db 721 -----HisHis----- 722
 QY 1451 AACATCTTGGCACAGGTGAGGGTCCAGAGGCACAGAACTTGGTGTCTGTATCAAGAC 1392
 Db 723 -----Cys--GlyGluProGluAspCysLysGluLeuGlyLe----- 734
 QY 1391 ACACAGGGTGGGACTGTAGCAGGTTC-----TGGAAATGTCC 1353
 Db 1041 nGlyLeuLeuGluProThrAlaTyLeuThrAla---HisSer-TrpArgLeuAsnProL 1060

735 -----CysValAlaGlyCysAsnCysProProGlyLeuLeuTrpAspLeuG 750
 QY 1352 AA--TTATGAGTACTTCCAGTAAAGTTTATGTGTAGA----- 1317
 Db 750 ludGlyGlnCysValProProSerMetCysHis-CysGlnPheGlyHisArgTyThr 769
 1316 -----GTTAGAAATAGAAAGCATGTCATCTGAACCTGAGTGGTTATGG 1272
 Db 770 IleAsnThrThrThrValArgAspCysSerHisCysileCysGlnGluArgGlyLeuTrp 789
 QY 1271 AAT-----AGATGCTGTATGGGGCCTTCTGTTGAGTGTAGAAA 1236
 Db 790 AsnCysThrAlaHisCysProArgGlnTrp-----AlaLeuCysProArgGlu 806
 1235 TTGTATATATGTCTGTGAATCTGGCAAGCTTCCCATGAGACATTTTGTGTAGCTAGA 1176
 Db 807 Leu-IleTyVal-----ProGlyAlaCysLeuLeuThrCysAs 819
 QY 1175 TAAT-----GTACTTCCCTG 1161
 Db 819 pSerProArgAlaAsnHisSerCysTrpAlaGlySerThrAspGlyCysValCysProPr 839
 1160 CACAAAGTACATAGCAATGACCAATGATGATGACGAGAATATATCTCCCACTGCTGT 1101
 Db 839 oGlyThrValLeuLeuAspLysHisCysValSer----- 850
 QY 1100 ACCTGAATTTGT 1052
 Db 851 -ProAspLeu-----CysProCysArgHisAsnGlyGlnTyPr 864
 1051 -----ACCATATTGAGATGATGATAAATGTTCTTTGTGTGTGTGTGTGTGTGTGTGT 1002
 Db 864 oProAsnAlaThrIle-----GlnGluAspCysAsnIleCys-----Va 877
 QY 1001 ATGCCAGAGACAGGTGGGAGGCATCCCTGGTCTCTGACAGTACAGTATGCTAT--TGG 945
 Db 877 lCysGln-----Gly--GlnArgTrpHisCysThrGlyGlnArgCysSerGlyTrp 893
 QY 944 ACACATAGCTGCCAAGAC----- 927
 Db 894 -----CysGlnAlaSerGlyAlaProHisTyValThrPheAspGlyLeuValPhe 910
 QY 927 ----- 927
 Db 911 ThrPheProGlyAlaCysGluTyriLeuValArgGluAlaGlyGlyArgPheSerVal 930
 QY 926 AGTATCCAAAGATTGTCACCTGCTGGAGTCCCTCCAGTGCCTCCACTGGGACACTTATGCTG 867
 Db 931 SerValGlnAsnLeuProCysGlyAlaSerGlyLeuThrCysThrLysAlaLeuValVal 950
 QY 866 AGGTTTGTCTCTGTTGAGTAGTTACTCTGTGACATGGACATGATGATTTTTCAGG 807
 Db 951 Arg-----LeuAspSerThrValValHisMetLeuArg 961
 QY 806 GGTCTCTGAAATGCTGT-----GTGTGTGTCAGAGGA 774
 Db 962 GlyGlnAlaValThrValAsnGlyValSerIleArgLeuProLysValTyThrGlyPro 981
 QY 773 GGCTTAGCCCCAGAAAGAACACCTCTGCTCTCTGCTAATTAACAG----- 725
 Db 982 GlyLeuSerLeuHisHisAlaGlyLeuPheLeuLeu-LeuThrThrArgLeuGlyLeuTh 1001
 QY 724 -----GCAGTGTGCTCTCT--TCTCCA----- 704
 Db 1001 rLeuLeuTrpAspGlyGlyThrArgValLeuValGlnLeuSerProHisPheHisGlyAr 1021
 QY 703 -----TCTGAGAAGCTTGGGTTGAGCAGACATCCAGA-- 671
 Db 1021 gValAlaGlyLeuCysGlyAsnPheAspSerAspAlaSerAsnAspLeuArgSerArgGl 1041
 QY 670 -GGCTCTCTTCATCT 612
 Db 1041 nGlyLeuLeuGluProThrAlaTyLeuThrAla---HisSer-TrpArgLeuAsnProL 1060

QY 897 CAGGACACAGTGGCAAAATCTTTGGTACTGTTCTGGCAGCTATGTTGTCACATGCAATG 956
 Db 138 Phe-----ArgGlnSerAlaLeuProPheHisLeuAlaArgAlaHisProGluLeu 155
 QY 957 CTCTTACTGACAGCAGCAGTCCCTCCACCTGTTCTGGCATACACCCACATGCAAG 1016
 Db 156 IleThrValThrAlaProSer-----ProSerThrLeuThrHisCysThrGlnCysPro 173
 QY 1017 CACAAAGACATTTATCCATCATCTCAATATGGTCCCAAGTGTGTGCATGCAATGCA 1076
 Db 174 ArgAlaPheHisSerSerAlaGlyLeuArgAsnHisSerArg-----IleHisVal 190
 QY 1077 ACACACACACACAAATTCAGGTAGCAGG-----TACGTGGCCAGTAT 1121
 Db 191 ValProSerLeuSerAspProGlyThrGluAlaHisLeuCysGlyIleCysGlyLys--- 209
 QY 1122 ATTCTGCTCATCAATAGGTGCTATGCTGTATCTTTGTCAGGAGAGTACATTTATCTACA 1181
 Db 210 -----SerPheSerLysSerSerThrLeuThrArg 219
 QY 1182 GTCAAAAATGCTCAT-----GGAAAGCCTTG 1211
 Db 220 HisLeuGlnArgHisSerGlyGluLysProPheLysCysProGluCysGlyLysPhe 239
 QY 1212 CCAGATTACACACATATATACAAATTCCTAACCAAGGCCCCCATACACCATCTATT 1271
 Db 240 LeuGluSerAlaThrLeuValArg-----HisGlnArgThrHisThrGlyGluLys 256
 QY 1272 CCATAAACCATCAGTTACAGATGATGCTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1331
 Db 257 ProTyrAlaCysSerAspCysGlyArgCysPheSerGluSerSerThr----- 272
 QY 1332 TTTACTGGAAGTACTCATATATTGGCATTTCCAGAACCTGCTCAGTCCACCCCTGTGT 1391
 Db 273 -----LeuLeuArgHisGlnArgSerHisGlnGlyGluArgProHisValCys 288
 QY 1392 GTCTTG-----ATACAGACACCAAGTTCTGT- 1420
 Db 289 AlaThrCysGlyLysGlyPheGlyGlnArgTyrAspLeuValHisGlnArgSerHis 308
 QY 1421 ---GCCTCTGACCTCAGTGGCCAGATGTTTAAAGTGTGATGTTTCAAAATTCATT 1477
 Db 309 ThrGlyGluArgProPheProCysProGlnCys-----GlyArgGlyPheThr 324
 QY 1478 GAAAGCTCTTTCTTGTA-----ACTCATGACAAAGTCCGCTCT---CATTCG 1522
 Db 325 AspArgSerAspLeuThrLysHisLeuArgThrHisThrGlyGluLysProTyrHisCys 344
 QY 1523 CAC-----TGAGAGTGTGTTAATGTGATCCAGACCTCTCTGTGAACATTTACCCCGC 1576
 Db 345 GluLeuCysGlyLysArgPheThrCysIleSerAsnLeuAsnVal-----HisLeuArg 362
 QY 1577 AAACCACTCAGCA-----AAGTGCCTTTCTTCCAGCAAGAAACAAAGCTCTT 1624
 Db 363 -AsnHisAlaGlyHisLysProHisLysCys-----ProGluCysGlyLysSerPheSe 380
 QY 1625 GGTGTGACTGCTAGAAATATGAAGCCACTCATTTATGTCTAGTGGACTGCAACTGT 1684
 Db 380 rValalaSerLysLeuAlaLeuHisArgLysThrHisLeuGlyGluArgThrAlaGlu-C 400
 QY 1685 GTACC---TGTGCAATGTTTACAGATGGAAGGGGTGAGGAGATGCTACACCTGACTAGG 1741
 Db 400 yThrGluCysGlyLysPhe----- 406
 QY 1742 TATCTCTATATAACCAAGTTTCCAGCAGGAGGAAGTACACAATCATCAGTGCAGTC 1801
 Db 407 -----PheSerHisGlyArgSerLeuSerGlnHisGln--ArgSe 419
 QY 1802 TCACAGAGG 1811
 Db 419 rHisArgArg 422

RESULT 11

Q8CJ95 PRELIMINARY; PRT; 1381 AA.
 AC Q8CJ95;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Tenbin 2.
 GN TENC1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lo S.H.;
 RT "Molecular cloning and characterization of mouse tensin 2.";
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF424789; AAN32753.1; -;
 DR MGD; MGI:2367586; Tenc1.
 DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
 DR InterPro; IPR008973; C2 CALB.
 DR InterPro; IPR002219; DAG_PE-bind.
 DR InterPro; IPR006020; PTB_PID.
 DR InterPro; IPR000980; SH2.
 DR Pfam; PF00130; DAG_PE-bind; 1.
 DR Pfam; PF00017; SH2; 1.
 DR ProDom; PD000093; SH2; 1.
 DR SMART; SM00109; C1; 1.
 DR SMART; SM00252; SH2; 1.
 DR PROSITE; PS00479; DAG_PE_BIND_DOM_1; 1.
 DR PROSITE; PS50081; DAG_PE_BIND_DOM_2; 1.
 DR PROSITE; PS50001; SH2; 1.
 SQ SEQUENCE 1381 AA; 149969 MW; D1C41CA0F0017786 CRC64;

Alignment Scores:

Pred. No.:	0.0375	Length:	1381
Score:	114.50	Matches:	104
Percent Similarity:	33.26%	Conservative:	50
Best Local Similarity:	22.46%	Mismatches:	147
Query Match:	1.43%	Indels:	162
DB:	11	Gaps:	24

US-10-776-827-10 (1-4460) x Q8CJ95 (1-1381)

QY 2513 TGGTACTACCTCCCTCCAGGTGACTTTAGGCTGTGGAGGTGGCAATTTAGTGTG 2572
 Db 694 TrpSerSerGluValGlyLysProLeuLeuHisProValArgProGlyHisProLeuPro 713
 QY 2573 ACCCTTGACACGAGGTTTCTTAACAGATGACCCCTGTGAATCATATATTTAAACCTGCATAT 2632
 Db 714 LeuLeuValProAla-----Cys-GlyHisHis----- 722
 QY 2633 ATTTATAGCAGTCACATTTGCCCTCTCACCTATATATGTC---CATAACTGCCTAAG- 2688
 Db 723 -----His-AlaProMetProAspTyrGlyCysLeuLysProProLysV 737
 QY 2689 -----CACTCAGGCTCCCACTCATCAACCCCTTGCACAGAGAAGAGC 2734
 Db 737 alGlyGluGluGlyHisGluGlyCysSerTyrAlaValCysSerGluGlyArgTyrGlyH 757
 QY 2735 ACTCTGGTCTCTATCCCTTGTGCATAGAGAGTTGTTCATGGGCTCTGG----- 2787
 Db 757 isSerGly-----TyrProAlaLeuValThrTyrGlyTyrGlyGlyAlaV 772
 QY 2788 -----CTGTGCCCTTCACATAACAGAATA-----ACTTGCCATCTGCCTGCA----- 2829
 Db 772 alProSerTyrCysProAlaTyrGlyArgAlaProHisSerCysGlySerProSerGluG 792
 QY 2830 -----CCAAACCCAGGATGTGGAGACATCTCCCAACCACTGCCACTGCTCAC 2878
 Db 792 lyArgGlyTyrProSerProGly-----AlaHisSerProArgAlaGlySerValSerP 810

[illegible]


```

Db      372 TrpSerProCys----- 375
QY      769 TAGCCCCAGAGGAAACACCTCTGCTTCCTGCTAACTAAACAGGCGAGTGATGCTCT 710
Db      376 -----SerGlyThrCys-----Thr 380
QY      709 TCTCCATCTGAGAGAGCTGGGTTGAGCAGAGATCCAGAGGTCTTCTTCATCCTCTCTCT 650
Db      381 AspProAlaHisProAlaTirp-----ArgSerArgSerArgLeuCysLeuAlaAsn 397
QY      649 TCTTCTCTCTCTCTCTCTCTCT----- 629
Db      398 CysThrGlyGlyAlaAlaSerGlnGluArgProCysAsnLeuProSerCysThrGluLeu 417
QY      628 ---CTGGCTGTGCGCCCTCTGCTGCTGGTGGCTCCAGCAGC----- 590
Db      418 ProLeuCysProGlyProGlyCysGluAlaGlyAsnCysSerTrpThrAlaTirpAlaPro 437
QY      589 -----AATGCCAACAGATGAAGAGAGGTGTGATGATCACCACAGAGTGGCAGGC----- 542
Db      438 TrpGluProCysSerArgSerCysGlyValGlyGlnGlnArgAlaGlyAlaTyrHis 457
QY      541 CCAGGAGGACTCCAGGAGATGTGTGTGTAGATGCTCTCT----- 500
Db      458 ProProGlyProGlyGlyHisTrpCysProAspValLeuThrAlaTyrGlnGluArgArg 477
QY      499 -----GGTCTTTGGTCAAGATGGTGTGATGATCACCACAGAGTGGCAGGC----- 470
Db      478 PheCysAsnLeuArgAlaCysProValProGlyGlyTrpSerArgTrp----- 493
QY      469 AGCTGAAGAGGCACTGATCTCTCCATGCTGTGGTGTGATGATGAGGCCAGAGAACCT 410
Db      494 -----SerProTrp-SerTrpCysAsp-----ArgSerCys 503
QY      409 CTCGGCTTACAGCAGTCTCTGAGCTCAAGAGATGCAACCTTGACCACTTGTAATTAT 350
Db      503 sGlyGlyArgSer---LeuArgSerArgSerCys----- 514
QY      349 TCCTCTTCCAGCTTGGGTTTGGGATGCGCCACTGACAGCAGCCTCCACACAGCAGCAAA 290
Db      515 -----SerSerPro-ProProGlyAsnG 522
QY      289 ATGGCCAGCAGCTGGGGAGCCTCTTCCAGCTTACTCTCTGCTTGTGCTGGGAAGGT 110
Db      522 lyGlyAlaProCysValGlyGluArgHisAlaArgLeuCysAsnProThrProCysG 542
QY      229 ATCTCGCCAGAACTGCTTCTCTGGAGGTCACTTCAACGTAAGAAATCGTGAACA 170
Db      542 luGluGly-----Cys---ProAlaGly----- 548
QY      169 CCGCCCTGTGATTCAGATCCTGATCACCGCCAGTGTGATGCTGCTGGGAAGGT 110
Db      549 --MetGluValValSerCysAlaAsnArgCysProArgCysSerAspLeuGlnGluG 568
QY      109 CACTCAGCTGCTTTAGTCTGATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 62
Db      568 lyIleValCysGlnGluAspGln-----AlaCysGlnGlnGlyCysArgCysProGluG 586
QY      61 AATCTTCTGCTCGGAGGAGGCTTCTCTG-----CAGGCC 26
Db      586 lySerLeuGlnGlnAspGlyGlyCysValProLeuGlyHisCysGluCysThrAspAlaG 606
QY      25 AGGGAATCTCTGG 12
Db      606 lnglyHisSerTrp 610

```

RESULT 14

```

Q8SPM4
ID Q8SPM4 PRELIMINARY; PRT; 5146 AA.
AC Q8SPM4;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)

```

```

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE SCO-spondin.
GN SCO-SPONDIN.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Subcommisural organ;
RA Meinzel A.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Subcommisural organ;
RX MEDLINE=20465125; PubMed=11008217;
RA Gobron S.;
RT "Subcommisural organ/Reissner's fiber complex: characterization of
RT SCO-spondin, a glycoprotein with potent activity on neurite
RT outgrowth.";
RL Gila 32:177-191(2000).
DR EMBL; AJ416457; CAC94914.1; -.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0005179; F:hormone activity; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR InterPro; IPR001064; Crystallin.
DR InterPro; IPR006207; Cys knot.
DR InterPro; IPR000421; FA5B C_.
DR InterPro; IPR001545; Gly_HormoneB.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR009041; PMP_inhibitor.
DR InterPro; IPR002919; TIL_Cysrich.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR001007; VWF C.
DR InterPro; IPR001846; VWF_D.
DR Pfam; PF00007; Cys_knot; 1.
DR Pfam; PF00754; F5_F8_type_C; 1.
DR Pfam; PF00057; ldl_recept_a; 10.
DR Pfam; PF01826; TIL; 10.
DR Pfam; PF00090; tsp_1; 25.
DR Pfam; PF00093; vwg_1; 1.
DR Pfam; PF00094; vwg_3; 1.
DR PRINTS; PR00261; LDLRECEPTOR.
DR SMART; SM00041; CT; 1.
DR SMART; SM00231; FA58C; 1.
DR SMART; SM00068; GHB; 1.
DR SMART; SM00192; LDLA; 10.
DR SMART; SM00209; TSP1; 25.
DR SMART; SM00216; VMD; 3.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 1.
DR PROSITE; PS01225; CTCK_2; 1.
DR PROSITE; PS01285; FA58C_1; 1.
DR PROSITE; PS01286; FA58C_2; 1.
DR PROSITE; PS00022; FA58C_3; 1.
DR PROSITE; PS00261; GLYCO_HORMONE_BETA_1; 2.
DR PROSITE; PS01209; LDLRA_1; 7.
DR PROSITE; PS00068; LDLRA_2; 9.
DR PROSITE; PS00092; TSP1; 25.
DR PROSITE; PS01208; VWF_C; 1.
DR PROSITE; PS0184; VWF_C_2; 2.
SQ SEQUENCE 5146 AA; 543576 MW; 724C5FB8727E13DA CRC64;

```

Alignment Scores:

```

Pred. No.: 0.0588 Length: 5146
Score: 114.50 Matches: 97
Percent Similarity: 28.25% Conservative: 29
Best Local Similarity: 21.75% Mismatches: 125
Query Match: 1.43% Indels: 195
DB: 6 Gaps: 26

```

US-10-776-827-10 (1-4460) x Q8SPM4 (1-5146)

QY	229	ATCTGGCCGAGAACTGCTTTCTTGAGAGTCAAGTTCAACGTAGTAACAATACTCGCTGAACA	170
Db	4051	luGlugly-----Cys---ProAlagly-----	4057
QY	169	CCGCCCTGTCATTGCAGATCCTGATCACCGCCAGTTCAGTGTGTGTCCTTGGGAAGGT	110
Db	4058	--MetGluValValSerCysAlaAsnArgCysProArgArgCysSerAspLeuGlnGluG	4077
QY	109	CACTCAGCTGCTTGTAGTCTCAGTTCCTCATCTCCAAAATGGG-----GAAC	62
Db	4077	lyIleValCysGlnGluAaspGln-----AlaCysGlnGlnGlyCysArgCysProGluG	4095
QY	61	AATCTTCGTCGCCGGAAGGGTCTCTG-----CAGGCC	26
Db	4095	lySerLeuGlnGlnAaspGlyGlyCysValProLeuGlyHisCysGluCysThrAspAlaG	4115
QY	25	AGGGAATCTCCTGG	12
Db	4115	InGlyHisSerTrp	4119
RESULT 15			
QZ5J3	ID	QZ5J3 PRELIMINARY; PRT; 451 AA.	
AC	QZ5J3	01-OCT-2003 (TREMBLrel. 25, Created)	
DT	01-OCT-2003	(TREMBLrel. 25, Last sequence update)	
DT	01-OCT-2003	(TREMBLrel. 25, Last annotation update)	
DE	BNF1		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Carnivora; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	PubMed=12853144;		
RA	Wu I.; Moes M.A.;		
FT	"BNF-1, a novel gene encoding a putative extracellular matrix protein,		
RL	is overexpressed in tumor tissues."		
FL	Gene 311:105-110(2003).		
DR	EMBL; AY163868; AA031809.1; --		
SQ	SEQUENCE	451 AA; 49675 MW; CFBFOBFA7D58CFE3 CRC64;	
Alignment Scores:			
Pred. No.:	0.0326	Length:	451
Score:	113.50	Matches:	90
Percent Similarity:	33.16%	Conservative:	38
Best Local Similarity:	23.32%	Mismatches:	124
Query Match:	1.41%	Indels:	134
DB:	4	Gaps:	23
US-10-776-827-10 (1-4460) x QZ5J3 (1-451)			
QY	318	GTGGCATGCCCAACCNAAGCTGGAAGAGGAATAAATTACAAGTGTCAAAGTTGCATC	377
Db	79	ValHisCysProGlnPro-----ValThrGluProGlnGlnCysCys	92
QY	378	CTTTTGAGCTCAGGACCTCGTTGT---AAGCCGAGAGGGTCTCTGGC-----	422
Db	93	-----ProLysCysValGluProHisThrProSerGlyLeuArgAlaPro	107
QY	423	CCTAATCTAGCCAAGCAC-----CATGGAGAG-----	449
Db	108	ProLysSerCysGlnHisAsnGlyThrMetTyrlinHisGlyGluIlePheSerAlaHis	127
QY	450	-----AATCAGTGCCTTCTT-----CAG	467
Db	128	GluLeuPheProSerArgLeuProAsnGlnCysValLeuCysSerCystThrGluGln	147
QY	468	CTCTATCTGGGACACCATCTTGACCAAAACACAGAAGGCATCTCAACACCATCTGCT	527
Db	148	IleTyrcysGly-----LeuThrThrCysProGlu-----Pro	158
QY	528	GGGAGTCTCTCGGGCGTCCACTTTGGTGATCATCATCTCTTCATCTGTTGCCA	587

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: September 21, 2004, 15:07:49 ; Search time 127 Seconds
(without alignments)
6756.129 Million cell updates/sec

Title: US-10-776-827-10
Perfect score: 8026
Sequence: 1 cggggcagcaaccaggagat.....aattaaaggtactactagctc 4460

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 566732

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+n2p_model -DEV=xlp
-Q=/cgn2_1/USFTC_spool_p/US10776827/runat_21092004_133241_20790/app_query.fasta_1.4615
-DB=PIR_78 -QFMT=fastan -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOOPEXT=0 -LIST=45
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LISF=45
-DOCALIGN=200 -THR SCORE=spct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10776827@cgn 1 221 @runat_21092004_133241_20790 -NCPU=6 -ICPU=3
-NO_MMAPP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR 78:.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	123.5	1.5	319	1 RWRBB	T-cell receptor be
2	117.5	1.5	628	2 S01955	hypothetical prote
3	113	1.4	628	2 S19150	hypothetical prote
4	112.5	1.4	631	1 A36749	transcription fact
5	112	1.4	597	2 JQ0107	hypothetical 66K p
6	109	1.4	968	2 T45746	hypothetical prote
7	108	1.3	332	2 S05500	Ig alpha-1 chain C
8	107	1.3	628	2 JQ0110	hypothetical 69K p
9	107	1.3	3020	2 A43932	mucin 2 precursor,
10	106.5	1.3	2843	1 RBHUAP	adenomatous polypo
11	105	1.3	630	1 A46149	transcription fact
12	104	1.3	1184	2 G01763	atrophin-1 - huma
13	103.5	1.3	330	2 S37595	mucin JUL10 - huma
14	103	1.3	353	1 A1HU	Ig alpha-1 chain C

15	102.5	1.3	649	2 T46500	hypothetical prote
16	102	1.3	962	2 S03818	carboxymethylcellu
17	101.5	1.3	628	1 A33333	transcription fact
18	101.5	1.3	769	1 JC1121	leukocyte adhesion
19	101.5	1.3	1184	2 S50832	atrophin-1 - human
20	101	1.3	1959	1 AGRT	agrin - rat
21	101	1.3	1964	2 T09059	otogelin - mouse
22	101	1.3	2910	2 T42214	otogelin - mouse
23	100.5	1.3	1210	2 I39410	AF-4 protein, spli
24	100.5	1.3	2825	2 T14271	Doc4 protein, stre
25	100	1.2	992	2 T38817	hypothetical prote
26	100	1.2	1257	2 S28764	neurocan precursor
27	100	1.2	3507	2 T34513	hypothetical prote
28	99.5	1.2	660	1 Q0BE3	BHLFI protein - hu
29	99	1.2	140	2 T46880	T-cell receptor be
30	99	1.2	753	2 JQ0532	OP protein - Kenne
31	99	1.2	2453	2 S60254	nuclear receptor c
32	98.5	1.2	542	2 S39608	transcription fact
33	98.5	1.2	543	2 S35047	mucin JUL7 - human
34	98.5	1.2	2395	1 S50820	surface protein ty
35	98.5	1.2	2845	2 T49505	adenomatous polypo
36	98	1.2	589	2 B38128	epithelin/granulin
37	98	1.2	814	1 T39627	nicotine dehydroge
38	97.5	1.2	1042	2 T48801	hypothetical prote
39	97.5	1.2	1213	2 A58198	serine/proline-ric
40	96.5	1.2	344	1 RWRTC2	T-cell surface gly
41	96.5	1.2	557	2 A55933	paxillin - human
42	96.5	1.2	628	1 A39262	transcription fact
43	96.5	1.2	1548	2 S34583	serine proteinase
44	96.5	1.2	1762	2 T03222	probable polykeid
45	96	1.2	333	2 G96780	unknown protein P9

ALIGNMENTS

RESULT 1

RWRBB
T-cell receptor beta chain precursor (ANA 11) - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C>Date: 30-Jun-1987 #sequence revision 30-Jun-1987 #text_change 16-Aug-1996
C:Accession: A94052; A94616; A02136
R:Angiolillo, A.L.; Lamoyi, E.; Bernstein, K.E.; Mage, R.G.
Proc. Natl. Acad. Sci. U.S.A. 82, 4498-4502, 1985
A>Title: Identification of genes for the constant region of rabbit T-cell receptor beta
A:Reference number: A94052; MUID:85242712; PMID:2989826
A:Accession: A94052
A:Molecule type: mRNA
A:Residues: 1-319 <ANG>
R:Mage, R.
submitted to the Protein Sequence Database, April 1987
A:Reference number: A94616
A:Contents: corrections to residues 79-83 and 104-105
A:Accession: A94616
A:Molecule type: mRNA
A:Residues: 79-83,104-105 <MAG>
C:Superfamily: immunoglobulin C region; immunoglobulin homology
C:Keywords: glycoprotein; heterotetramer; T-cell receptor; transmembrane protein
F:166-240/Domain: immunoglobulin homology <IMM>
F:292-313/Domain: transmembrane #status predicted <TM>
F:314-318/Domain: intracellular #status predicted <INT>
F:147,164,254,262/Binding site: carbohydrate (Asn) (covalent) #status predicted

Alignment Scores:	0.0151	Length:	319
Pred. No.:	123.50	Matches:	78
Score:	34.22%	Conservative:	38
Percent Similarity:	23.01%	Mismatches:	112
Best Local Similarity:	1.54%	Indels:	111
Query Match:	1	Gaps:	15

US-10-776-827-10 (1-4460) x RWRBB (1-319)

QY 954 ATGCTCTTACTGACAGCCAGCATGCTCCACCTGT----- 992

```

Db      24  LeuThrMetThrProGluHisAlaArgProProThrThrThrAlaMetLeuTyrAspGly 43
QY      993  ---CTTGGCATACCCACATGCAAGACGACAAAGACATTTATCCATACATCTCAATATG 1049
Db      44  PheLeuTrpProSerHisThrHisArgHisSerTyrLeuLeuHisProHisThrHisVal 63
QY      1050 GTTCCCAAGTGTGTGCATGCGTAAACACACACACACACAAATTCAGGTAGCAGGTAC 1109
Db      64  ---CysThrHisThrHisThrCys--ThrHisThrHisHis-- 76
QY      1110 GTGGGCAAGTATATCTGCTCATCAAAATGTCTATGCTATGCTATGCTATGCTAGGGAAGT 1169
Db      77  ---AlaSerThrHisValCysIleHisThrHisThrPheThrHisLeuCys-----11 93
QY      1170 ACATATCTACAGTCACAAAATGTCTCTGGAAGAGCTTGTCCAGATTCAGACACATAT 1229
Db      93  eHisThrLeuThrHisAlaLeuThrLeuThrCysAlaProThrArgThrTyrAlaHis-- 112
QY      1230 ATACAATTTCTTAACAGCAAGGCCCATACACCATCTATTCCATAAACCCCTCAGGTT 1289
Db      113  -----ThrArgAlaProThrHisValHisProHisLysProArgProAr 127
QY      1290 ACAGATGCATGCTTCTCTATTTCTAACTCTACACATAAACTTTTACTGGAAGTACTCATAT 1349
Db      127  gGlnLeuSerAlaAlaLeuLeuLeuProThrProLeuHisPheProGluAspLeuAlaAs 147
QY      1350 ATGGACATCTCCAGCAA-----CCTGCTACAGTC----- 1378
Db      147  nValSerAlaPro-GlnValValPheAspProSerGluAlaGluIleAsnLysThrG 167
QY      1379  -----CCCCACCTTGTGTGCTTGTATACAGAC----- 1405
Db      167  lnLysAlaThrLeuValCysLeuAlaLysAspPheTyrProAspHisValGluLeuSerT 187
QY      1406  -----ACACCAAGTTTCTGTGCTCTGACCCCTCACCTGTCCTGTCCTGTCCTGTCCTG 1445
Db      187  rpTnpValAsnGlyLysGluValHisAsnGlyValSerThrAspProGlnProTyrLysG 207
QY      1446 GATGTTTAAAGTGTGATGTTTCAAAATTCATTGAAAGCTCTTTTCTTTCTTTCTTTCTTT 1492
Db      207  lnAspProLysSerAspHisSerLysTyrCysLeuSerSerArgLeuArgValSerAlaA 227
QY      1493  --GTAACCTCATGACAAAGTCCGTCCTCAT-----TGCCACTGAGAGGTGTTTAAATGCA 1544
Db      227  lApheTrpHisAsnPro---ArgAsnHisPheArgCysGlnValGlnPhePheGlyLeuT 246
QY      1545 TCACAGACCTCTCTGTGAACACTTACCCCGCAACCACTACGCAAAAGTGCCTTTTCTCCA 1604
Db      246  hrAspAspAspGluTnpThrTyrAsnSerSerLysProIleThr----- 260
QY      1605 ACCAAGAACAAAGAGCTCTGTGGTGTGCTGTAGAAATTTATGGAAGCCCACTCATTTA 1664
Db      261  -----GlnAsnIleSerAlaHisThrArgG 269
QY      1665 TGTCACTGTGACTGCAACTGTGTACCTGTGCAATGTTTACAGATGGAAGGGTGAGGAGAT 1724
Db      269  lYArgAlaAspCysGly----- 274
QY      1725 GCTACACCTGTAGCTAGTATCTCTATATPAACAAAGTTTCCAGCAGGGAAGGAAGTACA 1784
Db      275  -----IleSerSerAlaSerTyrGlnGlnGly----- 283
QY      1785 CAATCATCAGTGCA-----GTCTCAGAGGAAGCAACACTG 1819
Db      284  --ValIleSerAlaThrValLeuTyrGluIleLeuLeuGlyLysAlaThrLeu 300

```

RESULT 2

S01955

hypotheical protein, 69K - turnip yellow mosaic virus

C:Species: turnip yellow mosaic virus, TYMV

C:Date: 21-Nov-1993 #sequence_revision 26-May-1995 #text_change 17-Mar-2000

C:Accession: S01955

R;Morch, M.D.; Boyer, J.C.; Haenni, A.L.
 Nucleic Acids Res. 16, 6157-6173, 1988
 A;Title: Overlapping open reading frames revealed by complete nucleotide sequencing of
 A;Reference number: S01955; MUID:88289359; PMID:3399388
 A;Accession: S01955
 A;Status: preliminary
 A;Molecule type: genomic RNA
 A;Residues: 1-628 <MOR>
 A;Cross-references: EMBL:X07441; NID:g62222; PIDN:CAA30321.1; PID:g62223
 A;Note: the authors translated the codon ACG for residue 459 as U
 C;Superfamily: hydroxyproline-rich glycoprotein

Alignment Scores:

Pred. No.:	0.0594	Length:	628
Score:	117.50	Matches:	101
Percent Similarity:	31.24%	Conservative:	38
Best Local Similarity:	22.70%	Mismatches:	141
Query Match:	1.46%	Indels:	165
DB:	2	Gaps:	20

US-10-776-827-10 (1-4460) x S01955 (1-628)

```

QY      2748 ATCCCTTGTGCATAGAGATTTGTATGGGGCCT-----CTGGCTGTGCCCTTC 2798
Db      40  LeuProMetValHisSerGluGlyThrSerThrProThrGlnLeuLeuArgHisProAsn 59
QY      2799 ACATACAGAAATTAACCTGCTGCTGCACCAACCCAGGATGTGGAAGACATCTCC 2858
Db      60  lIleTrpPheGlyAsnLeuProProProArgArgProGlnAspAsnArgAspPheSer 79
QY      2859 CCACAACCTCCACTGCTC----- 2876
Db      80  ProLeuHisProLeuValPheProGlyHisHisSerGlnLeuArgHisValHisGluThr 99
QY      2877 -----ACCAGCAAGCTGCCCTCTCTCTCCACTCTCAGCTCCCTCAGTATCCCTAGATGGATG 2930
Db      100  GlnGlnValGlnGlnThrCysProGlyLysLeuLysLeuSer----- 113
QY      2931 GCTGGGAGAGGTGGAGGCTGACAGTGAAGCTAGTGTGCATATGATCTAGGAGGGCG 2990
Db      114  ----- 115
QY      2991 GATACCCGGATCCGGACCATACAGTAACATGGT-----TTCCATGGCACTGCTTG 3044
Db      116  GluGluLeuProProAlaProGlnArgGlnHisSerLeuProLeuHis-IleThrArgPr 135
QY      3045 CTCG-----TTTGAATTAAAGACAGCAGTCAGTT----- 3072
Db      135  oSerArgPheProHisHisPheHisAlaArgArgProAspValLeuProSerValProAs 155
QY      3073 -----GTCAATTCATGACAGGCCCTCTATCT----- 3099
Db      155  pHisGlyProValLeuThrGluThrLysProArgThrSerValArgGlnProArgSerAl 175
QY      3100 -----CCAGGCACAAATGCTCCTGCTGCTCCTTAAT----- 3129
Db      175  aThrArgGlyProSerPheArgProIleLeuLeuProLysValValHisValHisAspAs 195
QY      3130 -----CCAATGGACTTGTCTCACCAGGGATGAACACCCAGAACTCACT----- 3177
Db      195  pProHisSerSerLeuArgProArgGlySerArgSerArgGlnLeuGlnProThrVa 215
QY      3178 -----TCTCAGTCACCTTCACAGCCCATGAC 3203
Db      215  lArgArgProLeuLeuAlaProAsnGlnPheHisSerProArgGlnProProLeuSe 235
QY      3204 TCAGAAAGAGCCAAACCCAGAAATGGGGCTCTCTTTC----- 3240
Db      235  rAspAspProGlyIleLeuGlyProArgProLeuAlaProHisSerThrArgAspProPr 255
QY      3241 -----CCCATCAGACTCCCTCTGCAAC-----CTTCTCTGGCG 3275
Db      255  oProArgProIleThrProGlyProSerAsnThrHisAspLeuArgProLeuSerVal 275

```

QY	3276	TAACTAGAGGAGTCCCACTGCAGATAGGCCCTAAACAGTTTGTAAATAAACAGGTGCA	3333
Db	275	uProArgThrSerProArgGly-----LeuLeuProAsnProArgArgHi	291
QY	3336	TGAAAGGAGCCTAAGGCCATTGTGATATCCACTCTCTCTTTCCACTTCCTTCTCATCT	3395
Db	291	sArgThrSerThr-GlyHis-----	297
QY	3396	TTTTTCCCATGTTTTATGTCTCTCTGATNTCCCTCTTGTGCTGCACACAGACCAGCC	3451
Db	298	-----lleProProThrThrSerArgProThrGlyP	309
QY	3452	--CCAGCCCTTTATCTCTCTCCATTTTCACCTTCCAGCTCTGTCCCTGAAGTCCAC	3509
Db	309	roProSerArgLeuGlnArgPro-ValHiLeuTyrGlnSer-SerProHisThrProAs	328
QY	3510	TGGCAACCCATGGGACCTCAGGACCAGAGACTGCTTCACTCATCTGCGG-----	3558
Db	328	nPheArgProSerSerIlearg--LysAspAlaLeuLeuGlnThrGlyProArgLeuGl	347
QY	3559	-----AGGTAAGTTTCACGGGGGACAAAAA	3584
Db	347	yHisLeuGluArgLeuGlyGlnProAlaAsnLeuArgThrSerGluArgSerProPro--	366
QY	3585	ATGATCTCTTAAGNAGAGGCTTCCTAGACAGACAGAGCTCCAGAAAGACATCCCTTAGG	3644
Db	367	-----ThrLyAargArgLeuProArgSerSerGlu---ProAsnArg--LeuProLys	382
QY	3645	CCTGGACTTCTGACAGACTTTAGCACAGGCTCCGACGGCAGCCAGAGAGGCGCTTTCCCC	3704
Db	383	Pro-----LeuProGluAlaThrLeuAlaProSerTyrArgHisA-GA-gProTyrPro	400
QY	3705	ATTGCTCTCTTCCCATTGCTCAATGATTCCAGTTTCTTTTCTTGGGGGAGACGAGG	3764
Db	401	LeuLeuProAsnProProAlaAa---LeuProSerIleAlaTyrThrSerSerArgGly	419
QY	3765	AGG 3767	
Db	420	Lys 420	
RESULT 3			
S19150			
hypothetical protein, 69k - turnip yellow mosaic virus			
C;Species: turnip yellow mosaic virus, TYMV			
C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 17-Mar-2000			
C;Accession: S19150			
R;Dreher, T.W.; Bransom, K.L.			
Plant Mol. Biol. 18, 403-406, 1992			
A;Title: Genomic RNA sequence of turnip yellow mosaic virus isolate TYMC, a cdna			
A;Reference number: S19150; MUID:92119261; PMID:1731998			
A;Accession: S19150			
A;Status: preliminary			
A;Molecule type: genomic RNA			
A;Residues: 1-628 <DRE>			
A;Cross-references: EMBL:X16378; NID:g62218; PIDN:CAA34414.1; PID:g62219			
C;Superfamily: hydroxyproline-rich glycoprotein			
Alignment Scores:			
Pred. NO.:	0.154	Length:	628
Score:	113.00	Matches:	103
Percent Similarity:	31.53%	Conservative:	37
Best Local Similarity:	23.20%	Mismatches:	139
Query Match:	1.41%	Indels:	165
DB:	2	Gaps:	22
US-10-776-827-10 (1-4460) x S19150 (1-628)			
QY	2748	ATCCCTTGTCACATAGAGCTTTGTCATGGGCGCT-----CTGGCTGTGCCCTTC	2798
Db	40	LeuProMetValHisSerGlnGlyThrSerThrProThrGlnLeuLeuArgHisProAsn	59
QY	2799	ACATAACAGAAATACTTGCCATCTGCTTGCACCAACCCAGGGATGTGGAAGACATCTCC	2858

Db	60	IleTrpPheGlyAsnLeuProProArgArgProGlnAspAsnArgAspPheSer	79
QY	2859	CCACAACTGCCATCTGTC	2876
Db	80	ProLeuHisProLeuValPheProGlyHisHisSerGlnLeuArgHisValHisGluThr	99
QY	2877	-----ACCAAGCAAGCTCCCTTCCTGCTCCACCTCTCAGTCCCCCTAGAAATGGATG	2930
Db	100	GlnGlnValGlnGlnThrCysProGlyLysLeuLysLeuSer	113
QY	2931	GCTGGGAGAGTGGAGGCTGACAGCTGACAGCTAGTGTGATATGATCTAGAGAGGCG	2990
Db	114	-----GlyAla 115	
QY	2991	GATCACCAGGATCCGGACCATACAAGTAACATGGT	3044
Db	116	GluGluLeuProProAlaProGlnArgGlnHisSerLeuProLeuHis-IleThrArgPr	135
QY	3045	CTCG-----TTTGAATTAAAGACAGCAGCTCAGTT	3072
Db	135	oSerArgPheProHisHisPheHisAlaArgProAspValLeuProSerValProAs	155
QY	3073	-----GTGATTCCATGACAAGGCTCTATCT	3099
Db	155	pHisGlyProValLeuAlaGluThrLysProArgThrSerValArgGlnProArgSerAl	175
QY	3100	-----CCAGGCCAAATGTCCTGCTGCTCCTTAAT	3129
Db	175	aThrArgGlyProSerPheArgProIleLeuLeuProLysValValHisAspAs	195
QY	3130	-----CCAATGGACTGTCTCACCCAGGATGAACACCCAGAACTCACT	3177
Db	195	pProProHisSerLeuArgProArgGlySerArgSerArgGlnLeuGlnProThrVa	215
QY	3178	-----TCTCAGTCATCTCCACAGCCGATGAC	3203
Db	215	lArgArgProLeuLeuAlaProAsnGlnPheHisSerProArgGlnProProLeuSe	235
QY	3204	TCAGAAGAGCCAAACCCAGAATGGGGCTCTCTTTTC	3240
Db	235	iAspAspProGlyIleLeuGlyProArgProLeuAlaProAsnSerThrArgAspProPr	255
QY	3241	-----CCCATCACAGACTCCCTGACAAC	3275
Db	255	oProArgProIleThrProGlyProSerAsnThrHisGlyLeuArgProLeuSerValLe	275
QY	3276	TAACTAGAGAGTCCAGTGCAGATAGGCCCTTAACGTTTGTAAATAACAGGTGCA	3335
Db	275	uProArgAlaSerProArgArgGly-----LeuLeuProAsnProArgArgHi	291
QY	3336	TGAAGAGGCTTAAGCCATTGTTGATACCACTCTCTCTTCCACTT	3389
Db	291	sArgThrSerThrGlyHisIleProProThrThrThrSerArgProThrGlyProProSe	311
QY	3390	TCATCTTTTTCCTCCATGTTTATGCTTCTCTGATTCCTCTCTCGCTGCACACACAG	3449
Db	311	rArgLeuGlnArgProVal-----HisLysTyrrGlnSerSe	323
QY	3450	CCCCAGCCCTTTATTCCTCTCCATTTTCACTCTCTTCCAGCTC	3492
Db	323	rPro-----HisThrProAsnPheArgProSerSerIleArgLysAspAlaLeuLe	340
QY	3493	-----TGTCCTCGAAGTCCACTGCCACTGCCACCCCATG	3521
Db	340	uGlnThrGlyProArgLeuGlyHisLeuGluCys-----LeuGlyGlnProAl	356
QY	3522	GGACCTCAGACACAGACTGCTTGACTCATCTGGGAGGGTAAAGTTTCACGGGGGACAAA	3581
Db	356	aAsnLeuArgThr-----SerGluArgSerProPr	366
QY	3582	AAATATGATCTCTAAAGAAGAGGCTTCTTAGACACAGCAGGCTCCAGAAACATTCOCCT	3641

```
Db 366 o-----ThrLysArgLeuProArgSerSerGlu---ProAsnArg--LeuPro 381
QY 3642 AGCCTGGACTTCTGACAGCTTTAGCCAGGCTCCGACGGCAGCAGAGGAGGCTTTC 3701
Db 382 LysPro-----LeuProGluAlaThrLeuAlaProSerTyrArgHisArgArgSerTyr 399
QY 3702 CCATTGCTCTTCCCATGCTGCTCAATGATTCCATGTTCTTTCTTGGGGGAGCA 3761
Db 400 ProPheLeuProAsnProProAlaAla---LeuProSerIleAlaTyrThrSerSerArg 418
QY 3762 GGGAGG 3767
Db 419 GlyLys 420

RESULT 4
A36749
transcription factor HNF-1A - human
N;Alternate names: hepatic nuclear factor 1-alpha; transcription factor 1, hepatic; tran
C;Species: Homo sapiens (man)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: A36749
R;Bach, I.; Galcheva-Gargova, Z.; Mattei, M.G.; Simon-Chazottes, D.; Guenet, J.L.; Cereg
Genomics 8; 155-164, 1990
A;Title: Cloning of human hepatic nuclear factor 1 (HNF1) and chromosomal localization o
A;Reference number: A36749; MUID:91184801; PMID:1707031
A;Accession: A36749
A;Molecule type: mRNA
A;Residues: 1-631 <BAC>
A;Cross-references: GB:M57732; NID:g184264; PIDN:AAA88077.1; PID:g184265
C;Genetics:
A;Gene: GDB:TCF1; HNF1; LFB1
A;Cross-references: GDB:125297; OMIM:142410; OMIM:600496
A;Map position: 12q24.3-12q24.3
C;Complex: homodimer; can also form heterodimers with, for example, HNF-1B
C;Function:
A;Description: transcription activator required for the expression of a number of liver-
A;Note: also expressed in some other tissues, where it may play other roles
C;Superfamily: transcription factor HNF-1; homeobox homology
C;Keywords: DNA binding; heterodimer; homeobox; homodimer; liver; nucleus; transcription
F;1-33/Region: dimerization
F;222-277/Domain: homeobox homology <HOX>

Alignment Scores:
Pred No.: 0.172 Length: 631
Score: 112.50 Matches: 103
Percent Similarity: 31.55% Conservative: 44
Best Local Similarity: 22.10% Mismatches: 180
Query Match: 1.40% Indels: 139
DB: 1 Gaps: 23

US-10-776-827-10 (1-4460) x A36749 (1-631)
QY 225 GAGATCCCTGACACGCAAGCAAGTAAGTAACGTGGAAGAGGCTCCCAAGCTGGCTG 284
Db 193 GluLeuPro-----ThrLysLysGlyArgArgAsnArgPheLysTrpGlyProAlaSerGln 211
QY 285 GCCATTTCGTGCTGTGTGTGAGGTGCTGTCAGTGGCATGCCCAACCCAAAGCTGAA 344
Db 212 GlnIleLeuPheGlnAlaTyrGlu-----ArgGlnLysAsnProSerLysGlu 227
QY 345 GAGCAATAAATACAAAGTGGTCAAGGTGTGATTCCTTTTGAGCTCAGGACCTGCTGTAAG 404
Db 228 GluArg--GluThrLeuValGluGluCys-----AsnArgAlaGluCysIle 242
QY 405 CCGAGAGGTTCTCTGGCCCTTAATCTAGCCAAAGCACCAGCAGAGATCAGTGCCTCTT 464
Db 243 GlnArgGlyValSer--ProSerGlnAlaGlnGlyLeuGlySerAsnLeuValThrGlu 261
QY 465 CAGCTCTATCTGGGACACCATCTTGACCAACACCAAGAGGATCTACACACCATCTG 524
Db 262 ValArgValTyrAsnTrpPheAlaAsnArgArgLysGluGluAlaPheArgHisLysLeu 281
QY 525 -----CCTGGAGTCTCTCTGGGCTGCCACTCTTGGT 557

282 AlaMetAspThrTyrSerGlyProProGlyProGlyProGlyProGlyProAla---LeuPro 300
558 GATCATCACACTCCTCTTTCATCTGTGTCATCTGCTGAGCCGCCAGGAGGGG 617
301 AlaHisSerSerPro---GlyLeuProProProAlaLeuSerProSerLysValHisGly 319
618 -----CCAGCAGCCAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGATGA 662
320 ValArgTyrGlyGlnProAlaThrSerGluThrAlaGluValProSerSerSerGly--- 338
663 AGAAGACCTCTCGATCTCTGCTCAACC-----CAAGTCTCTCAGATGAGAGAG 713
339 ---GlyProLeuValThrValSerThrProLeuHisGlnValSerProThrGlyLeuGlu 357
714 ACCATCAGCTGCTGTTAGTTAGGAGGAGGAGAGGTTCTTCTTCTGGGCTAAGCC 773
358 Pro---SerHisSerLeuLeuSerThrGluAlaLysLeuValSerAlaAlaGly----- 374
774 TCCTCTGACCACACACAGACATTTTCAGGAACCCCTGAAATATGCACTATGTCATGTC 833
375 -----GlyProLeuProProValSerThrLeuThrAlaLeu 386
834 CACAGA-----GTAACCTACTCAACCAAGAGAAACAACTCAGACT 872
387 HisSerLeuGluGlnThrSerProGlyLeuAsnGlnGlnProGlnAsnLeuIleMetAla 406
873 AGTGTCCAGTGGAGGAGGAGTCCAGGACCA----- 905
407 SerLeuPro---GlyValMetThrIleGlyProGlyGluProAlaSerLeuGlyProThr 425
905 ----- 905
426 PheThrAsnThrGlyAlaSerThrLeuValIleGlyLeuAlaSerThrGlnAlaGlnSer 445
906 -----COTGGACAATCTCTGGATCTCTCTGGAGCT----- 938
446 ValProValIleAsnSerMetGlySerSerLeuThrLeuGlnProValGlnPheSer 465
939 -----ATGTGTCCCAATAGCAATGCTCTTACT 965
466 GlnProLeuHisProSerTyrGlnGlnProLeuMetProProValGlnSerHisValThr 485
966 GCAGACCCAGGATGCTCCACCTGTCTGTCATACCCACATGCA-----AAGCAC 1019
486 GlnSerProPheMetAlaThrMetAlaGlnLeuGlnSerProHisAlaLeuTyrSerHis 505
1020 AAGAACATTTATCC-----ATACATCTCAATATGTTCCCAAGTGTGTCATGTCAC 1073
506 LysProGluValAlaGlnTyrThrHisThrGlyLeuLeuProGlnThr-----MetLeu 523
1074 GTAACACACACACACACAAATTCAGGTAGCAGGTACGTGGCAAGTATATTCTGCTCATC 1133
524 IleThrAspThr----- 527
1134 AAATGGTCAATTGGCTATGTACTTTTGGCAGGGAAGTACATATTCTACAGTCACAAAAATG 1193
528 -----ThrAsnLeuSerAlaLeuAlaSerLeu 536
1194 TCTCATGGGAAAGCCTTGCAGATTCAGACACATATATACAAATTCCTACACGACAGGC 1253
537 ThrProThrLysGlnValPheThrSerAspThr-----GluAlaSerSerGluSerGly 554
1254 CCCCATACCATCTATTCCATAAACCCACTCAGGTTACAGATGCTTCTTCTTATTCT 1313
555 LeuHisThrProAlaSerGlnAlaThrLeuHisValProSerGlnAspProAlaGly 574
1314 -----AACTCTACACATAAACTTTTACTGGAAGTACTCATATAATTGGACATTC 1361
575 IleGlnHisLeuGlnProAlaHisArgLeuSerAlaSerProThrValSerSerSer 594
1362 AGCAACTGTCTACAGTCC 1379
```

```

Db      595 LeuValLeuTyrGlnSer 600
RESULT 5
JQ0107
hypothetical 66K protein - Ononis yellow mosaic virus
C:Species: Ononis yellow mosaic virus
A:Note: host Pisum sativum cv. Early massey (pea)
C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 08-Oct-1999
C:Accession: JQ0107
R:Ping, S.W.; Keese, P.; Gibbs, A.
Virology 172, 555-563, 1989
A:Title: Nucleotide sequence of the ononis yellow mosaic tymovirus genome.
A:Reference number: JQ0106; MUID:90021186; PMID:2800337
A:Accession: JQ0107
A:Molecule type: genomic RNA
A:Residues: 1-597 <DIN>
A:Cross-references: GB:J04375; NID:g332572; PIDN:AAA46795.1; PID:g332573

Alignment Scores:
Pred. No.:      0.189      Length:      597
Score:          112.00     Matches:      96
Percent Similarity: 33.25%  Conservative: 45
Best Local Similarity: 22.64% Mismatches:    121
Query Match:     1.40%     Indels:      163
DB:             2         Gaps:        24

US-10-776-827-10 (1-4460) x JQ0107 (1-597)
QY      2656 CCTCTCACCTATATGCGCCATAAAGTGCCTAAGCACTCAGGCCTC----- 2700
Db      187 ProLysProLeuHisLeuHisAsnProArgGlnHisSerSerLeuArgSer 206
QY      2701 -----CCACTCATCAACCCCTTTGACCAGAGAAAGAAC 2736
Db      207 ProCysArgLysLeuArgProThrProArgCysAsnGlnLeuAlaGlnGlnHis 226
QY      2737 -----TCTGGTTCTATATCCCTTCTCATATA 2763
Db      227 ProLeuProSerSerLysProLeuSerLeuGlnAlaGlyLeuGlyProCysProLeu 246
QY      2764 GAGAGTTTGTGATGGGCGCTCTGGCTGTGCCTTCCATCAACAGATAAATGCGCATCTG 2823
Db      247 -----ProProHisAsnLys----- 251
QY      2824 CTGTGACCAAAACCCAGGATGTGGAGACATCT-----CCCAACAAGTGCAC 2871
Db      252 ArgSerProSerProAlaValIleArgLysThrAlaGlyLeuProHisProLysLeu 271
QY      2872 TGCTCACCAGCAAGTGCCTCTCTGCTCTCCACTCTCAGTCCCTCAGATGATGGATGG 2931
Db      272 ProProSerSerArgGlyHisLeuProSerSerSerSer----- 286
QY      2932 CTGGGAGAGCTGGAGCTGACAGCTGAGACGTAGTGTCTAGATATGATCTAGGAGGCGG 2991
Db      287 -----SerProArgSerAsnArgGlyValArg 295
QY      2992 ATCACCAGGATCCGGGACCATACAAGTAACATGTTTCCATG---GCAACTGTGTC--- 3045
Db      296 CysSer---ValHisLeuHisLysSerArgSerAsnSerGlnAspLeuArgSerCysArg 314
QY      3046 -----TCGTTTGNATTAAGACAGCAGTCACTGTTGCTATTGCCATGACAGCCCTCTA--- 3096
Db      315 ValArgSerAsnSerLeuGlnGlnThrProLeuLeuMetGlyHisPheLysSerLeuGly 334
QY      3097 ---TCTCCAGCAATGTCCTGCTCTCTCTAATCCA-----ATG 3135
Db      335 GlnSerProAsnLeuArgSerPheGluArgProArgProThrArgArgSerLeuArgLeu 354
QY      3136 GACTTGCTCTCACCCAGGATGAACACCCAGCAAACTCACTTCTCAG----- 3183
Db      355 LeuProLeuSerProGlnLysValProThrValHisValProThrHisGlnGlnSerGly 374
QY      3184 -----TCACTTCCACGCCGATGACTCAGAGAGCCAAACCCAGATGGGGCC 3231

```

```

Db      375 HisLysGlyProSerLeuProArgProHisSerProSerArgGlnThrHisAla--- 393
QY      3232 TCTCTTTTCCCATCACAGACTCCCTGACAACCTTTCTCGCGTAACCTAGAGAGTCC 3291
Db      394 -----ArgLeuPro 396
QY      3292 AGTGCAGATAGCCCTAAACGTTTGTAAATAAACAGGTG-----CATGAAGGAGC 3345
Db      397 His-----SerLysArgValSerLeuProAsnSerValLeuHisAspArgPro 413
QY      3346 CTAAGGCCCATGTGTATATCCACTCTCTCTTCTTCCACTT-----CCTTCTCATCTT 3396
Db      414 LysArgProIleHisPheGlySer-----PheProIleAsnValAlaProSerHisLeu 431
QY      3397 TTTCTCCATGTTTATGTCTTCTGTGATCCCTTCTGTGCTGCTGACAGACCCAGC 3456
Db      432 Leu-----ProArg 434
QY      3457 CCTTATTTCTCTCCATTTTTCACCTCTCCAGCCTCTGTCCCTGAAGTCCACT---GGC 3513
Db      435 LysLeuTrpSerArgAlaSerSerProProThrCysSerProThrThrSerAsnHisGly 454
QY      3514 AACCCATGGAC---CTCAGGACCAAGACTGCTTGACTCATCTGGGAGGAGTAAGTTCA 3570
Db      455 HisProGluGluAlaLeuArg----- 461
QY      3571 CGGGGGACAAAAAATGATTTCTTAAAGAGAGGCTTCTTAGACACACAGGC----- 3623
Db      462 -----PheLeuProLysAsnLeuProGlnHisCysGlnMetAlaLeu 475
QY      3624 -----TCCAGAAAGACATCCCTTAGCCCTGGACTTCTGACGACCTTTAGC 3668
Db      476 MetGluAsnTyrCysSerHisPheSerSerPro-----SerSerSerVal 490
QY      3669 CAGGCTCCGAGCGGACGAGGAGGCGCTTTCCCATCTGCTCTTCCCATTTGCTCAA 3728
Db      491 SerPheProGluAspHisGlnSerSerLeuProProIleSer-----ThrArg 506
QY      3729 TGGATTCCCATGT 3740
Db      507 TrpValGlnCys 510

RESULT 6
T45746
hypothetical protein F24M12.190 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C:Accession: T45746
R:Vitale, D.; Liguori, R.; Flores, M.; Argiriou, A.; De Simone, V.; Mewes, H.W.; Lemcke,
submitted to the Protein Sequence Database, December 1999
A:Reference number: Z23012
A:Accession: T45746
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-968 <VIT>
A:Cross-references: EMBL:AL132980
A:Experimental source: cultivar Columbia; BAC clone F24M12
C:Genetics:
A:Map position: 3
A:Introns: 69/1; 99/1; 132/3; 170/3; 198/1; 220/3; 242/3; 284/2; 396/3; 441/3; 514/1; 81
A:Note: F24M12.190

Alignment Scores:
Pred. No.:      0.384      Length:      968
Score:          109.00     Matches:      99
Percent Similarity: 36.65%  Conservative: 74
Best Local Similarity: 20.97% Mismatches:    150
Query Match:     1.36%     Indels:      149
DB:             2         Gaps:        23

US-10-776-827-10 (1-4460) x T45746 (1-968)

```

QY	47	CGGACCGAAGAGATGTTTCCCATTTTGGAGATGAAGAACTGAGACTCAAGCAGCTGCA	106
		::: :::: ::::	
Db	352	SerAspLysAlaLeuValArgHisLeuGlnArgGluLeuAlaLysLeuGluSerGlu---	370
		::: :::: ::::	
QY	107	GTGACCTTCCCAAG--	148
Db	371	LeuSerSerProArgGlnAlaLeuValSerAspThrThrAlaLeuLysGluLys	390
QY	149	GATCTTGAAATGCACGGGGGTGTTTCAGCGATTGTTTACTAGCTTGAACGTGACCTCCAG	208
		::: ::::	
Db	391	AspLeuGlnIleGlu--LysLeuAsnLysGluValPheGlnLeuAlaGlnIleGluGlu	409
		::: ::::	
QY	209	GAAAGCAGTTCT-	232
Db	410	ArgAlaTyrSerArgIleGluAspLeuGlnIleIleGlyLeuAlaProGlnGlnGlu	429
QY	233	---CTGACAAACCAAGCAAGAAAGTAACGT-	274
		::: ::::	
Db	430	IleLeuSerThrAspSerGluGlnThrAsnThrAsnValValLeuGlyArgGlnTyrPro	449
		::: ::::	
QY	275	AAGCTGGCTGGCCCATTTTCTCTCTGCTGTGTGAGGTGCTGTCAAGTGGCATGCCAAACCC	334
		::: ::::	
Db	450	LysLeu---ArgValArgSer-SerTrpGluSerLeuAsnIleThrProGluSerPr	467
		::: ::::	
QY	335	AAAGCTGGAAGAGGAATAAATCAAGTGGTCAAGGTTCATCTCTTTTGAGC-	386
		::: ::::	
Db	467	O-----LeuSerAlaGlnAlaSerIleMetIleSerProGlnSe	480
		::: ::::	
QY	386	-----	386
		::: ::::	
Db	480	rThrGluHisGlySerAspGluAsnValPheGlnLeuSerAspLeuArgLeuAsnSerG	500
		::: ::::	
QY	387	---TCAGACCTGCTGTGAAGCCGAGAGGGTCTCTGGCCCTTAATCTAGCCAAGCACCA	442
		::: ::::	
Db	500	YAlaSerSerProAla--GlnHisLeuAlaPheValThrProAspGlnSerArgValG	519
		::: ::::	
QY	443	TGGAGAGAATCAGTGCCTTCTTTCAGCTCTATCTGGACACCATCTTTGACCAC-	494
		::: ::::	
Db	519	nGlyGluArgLeu-----HisGluMetAspGluProSerG	531
		::: ::::	
QY	495	-----ACACAAAGAGGCATCTACAAACCATCTCGCTGGG--	530
		::: ::::	
Db	531	uValAspSerGluAspThrCysThrGluLeuGlnCysIleGluThrGluSerProGlyI	551
		::: ::::	
QY	531	---AGTCCTCTGGGCGCTGCCACTCTTGCTGATCATCACATCTCTTCTTCTGTTGCC	586
		::: ::::	
Db	551	eIleMetTyrProGluProAsnIleLeuProAspArgCysLysAlaValSerAlaLeuPr	571
		::: ::::	
QY	587	ATTGCTGC-----TGGAGCCCAACGAGGAGGGGCCAGCAGCCAG	628
		::: ::::	
Db	571	oleu-CysGluProGluSerLysAsnSerArgProThrGluThrAlaGluGluLysG	591
		::: ::::	
QY	629	AGAAGAAAAGAGAAGAAGAAGATGAAGAAGACCTCTCGATCTCTGCTCAAC	688
		::: ::::	
Db	591	luGluLysGluGluThrGluGluLysGluGluGluGluGluGluGluValLysGluVal	611
		::: ::::	
QY	689	CCAAGCTTCTCCAGATGGAGAAGACCATCACTGCCTGTTAGTTAGCAGGAGACAGA	748
		::: ::::	
Db	611	erSerValSerIleGlnThrLysGluLysSerGlyProIleLysVal-----SerProA	629
		::: ::::	
QY	749	GTGTTTCTCTTCT- ---GGGGCTAAGCCTCTTCTGACCACACACAGACATTTTCAG	802
		::: ::::	
Db	629	rgCysValLeuSerLeuThrAspGluSerPheProAspGluSerSerAsnLeu-LysArg	648
		::: ::::	
QY	803	AACCCCTGAAAT---AATGCATATGTCATGTCCACAGAGTAACCTACTCAACCAAGGAA	859
		::: ::::	
Db	649	AspProThrHisGlnAspPheValThrProSerProGluLysLeuTyrAlaTrpHisLeu	668
		::: ::::	
QY	860	CAAACCTCAGACTAAGTGTCCAGTGGAGGCGAGTCCCGAGGACACAGCTGGCAATTC	919
		::: ::::	
Db	669	Glu-----SerAsnGlyGlnThrAlaGlyGlyThrGlyPheThrArgSer	683
		::: ::::	
QY	920	GGATACTGCTTGGCAGCTATGTGTCCAATAGC-----	952
		::: ::::	

Db	684	ArgSerCysGlyAlaSerPheValSerSerPheSerLeuSerGluArgAspAla	703
Qy	953	AATGCTCCTTACTGC	967
Db	704	AsnThrProCysTrpTyrGlnAsnGluAlaGluSerAsnLeuLysProSerAsn	723
Qy	968	-----AGACCCAGCATGCTCCCACTGTCTCTGGCATACCCACATGC	1012
Db	724	SerLysArgProProLeuProLysHisIleSerArgMetSerMet	741
Qy	1013	AAAGCACAAGAACATTATCCATACA	1039
Db	742	TrpPheGluLysAspPheAsnHisThrGlnArgMetProAlaGlyLeuAspGlyValAsn	761
Qy	1040	-----TCTCAATATGGTTCCCAAGTGTGTGCATGCACGTAACACACACACA	1087
Db	762	MetIleLysSerSerProAsnGlySerGlnVal	777
Qy	1088	CACAAATTCAGTAGTCAGGTACGTGGCCAA	1117
Db	778	HisValTyrAlaArgGlnThrSerGlyArg	787
RESULT 7			
S05500			
Ig alpha-1 chain C region - gorilla (fragment)			
C:Species: Gorilla gorilla (gorilla)			
C:Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 23-Jul-1999			
R:Accession: S05500			
R:Kawamura, S.; Omoto, K.; Ueda, S.			
Nucleic Acids Res. 17, 6732, 1989			
A:Title: Nucleotide sequence of the gorilla immunoglobulin alpha 1 gene.			
A:Reference number: S05500; MUID:89386006; PMID:2506527			
A:Accession: S05500			
A>Status: translation not shown			
A:Molecule type: DNA			
A:Residues: 1-352 <XAM>			
A:Cross-references: EMBL:X15045; NID:g22900; PIDN:CAA33147.1; PID:g22901			
C:Genetics:			
A:Introns: 102/1; 222/1			
C:Superfamily: immunoglobulin C region; immunoglobulin homology			
C:Keywords: immunoglobulin			
F:242-314/Domain: immunoglobulin homology <IMM>			
Alignment Scores:			
Pred. No.:	0.408	Length:	352
Score:	108.00	Matches:	75
Percent Similarity:	28.37%	Conservative:	24
Best Local Similarity:	21.49%	Mismatches:	89
Query Match:	1.35%	Indels:	161
DB:	2	Gaps:	16
US-10-776-827-10 (1-4460) x S05500 (1-352)			
Qy	669	CCTCTGGATCTCTGCTCAACCCAGCTTCTCCAGATGGAGAGACCATCAGTCCTGT	728
Db	9	ProLeuSerLeuCysSerThrGln-----ProAspGly---AspValValValAlaCys	25
Qy	729	TTAGTTAGGACGAGAACGAGGTTTCCT	770
Db	26	LeuValGln-----GlyPhePheProGlnGluProLeuSerValThrTrpSerGlu	42
Qy	771	GCC-----TCCTTCGACACACACAGACATTTTCAGGAACCCCT	809
Db	43	SerGlyGlnGlyValThrAlaArgAsnPheProProSerGlnAspAlaSerGlyAsp	61
Qy	810	GAATAATGCACTATGTCTCATGTCACAGAGTAAGTACTACCAACCAAGGAACAAACCTCAG	869
Db	62	-----LeuTyrThrThrSerSerGlnLeuThrLeuProAla	73
Qy	870	ACTAAGTGTCCAGTGGGCGCAGTCCACGGACCACTTCCTGGATACTGTC	929
Db	74	ThrGlnCysProAspGly-----	79

QY 930 TTGGCAGCTATGTGTCACATAGCAATGCTCTTACTGACAGCCAGGATGCTCCACC 989
 Db 79 ----- 79
 QY 990 TGTCTCTGGCATACCCACATGCAAGCAAAAGCAATTATCCATACATCTCAATATG 1049
 Db 79 ----- 79
 QY 1050 GTTCCCAAGTGTGACATGACGTAACACACACACACAAATTCAGGTAGCAGGTAC 1109
 Db 80 -----LysSerValThrCysHisValAsnHis---TyrThrAsnProSer----- 93
 QY 1110 GTGGGCAAGTATATCTGCTCATCAAAATGGTCAATTGGCTATGTACTTTGTGCAGGGAAGT 1169
 Db 93 ----- 93
 QY 1170 ACATTATCTACAGTCACAAAATGTCTCATGGGAAGCCTTGGCAGATTCCAGACACATAT 1229
 Db 94 -----GlnAspValThrValProCysArgValProSerThrProProThr----- 108
 QY 1230 ATACAAATTTCTTAACCAACGAGGCCCATACACCATCTATTCCATAAACCACTCAGGTT 1289
 Db 109 -----ProSerProSerThrProProThrProSerProPro----- 120
 QY 1290 ACAGATGATGCTTTCCTATTCTTAACCTCTACACATAAACTTTTACTGGAAGTACTCATA 1349
 Db 121 -----CysCysHisPro---ArgLeuSerLeuHisArgProAlaLeuGluAspLeuLeu 137
 QY 1350 ATTGGACATTCGACCAACTG----- 1370
 Db 138 LeuGlySerGluAlaAsnLeuThrCysThrLeuThrGlyLeuArgAspAlaSerGlyVal 157
 QY 1371 -----CTACAGTCCCCACCCCTTGTGTGTC 1394
 Db 158 ThrPheThrTrpThrProSerSerGlyLysSerAlaValGluGlyProProGluArgAsp 177
 QY 1395 TTGATACACACACACCAAGTTTCTGTGCTCTGACCCCTCACCTGTGCCAAGATGTTAA 1454
 Db 178 LeuCysGlyCysTyrSerValSerValLeuPro---GlyCysAlaGluProTirpAs 196
 QY 1455 AGTGTGATGGTTCAAAATTCATGAAGCTCTTTCTTGAATCTACTCATGACAAGTCCGTC 1514
 Db 196 n-----HisGlyLysThrPheThrCysThrAlaAlaLysrProGluSe 210
 QY 1515 CTCATTGCCACTCAGAGGTGTTTAATGTATCCA-----AGACCTCTCTGTGA 1562
 Db 210 rLysThrProLeuThrAlaThrLeuSerLysSerGlyAsnMetPheArgProGluValHi 230
 QY 1563 AACATTACCCCGCAACCAACTCAG 1587
 Db 230 sLeuLeuProProProSerGluGlu 238
 RESULT 8
 JQ0110
 Hypothesis: 69K protein - turnip yellow mosaic virus
 C:Species: turnip yellow mosaic virus, TYMV
 C>Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 19-May-2000
 R:Accession: JQ0110
 R:Keese, P.; Mackenzie, A.; Gibbs, A.
 Virology 172, 536-546, 1989
 A>Title: Nucleotide sequence of the genome of an Australian isolate of turnip yellow mosaic virus
 A:Reference number: JQ0109; MUID:90021184; PMID:2800335
 A:Accession: JQ0110
 A:Molecule type: genomic RNA
 A:Residues: 1-628 <KEE>
 A:Cross-references: GB:J04373; NID:g332244; PID:AAA46591.1; PID:g332245
 A:Experimental source: Australian isolate
 C:Superfamily: hydroxyproline-rich glycoprotein

Alignment Scores: 0.55 Length: 628
 Pred. No.: 107.00 Matches: 96
 Score:

Percent Similarity: 30.57% Conservative: 37
 Best Local Similarity: 22.07% Mismatches: 155
 Query Match: 1.33% Indels: 147
 DB: 2 Gaps: 17
 US-10-776-827-10 (1-4460) x JQ0110 (1-628)
 QY 2748 ATCCCTTGTGCATAGAGAGTTTGTTCATGGGCGCT-----CTGGCTGTGCCCTTC 2798
 Db 40 LeuProMetValHisSerGluGlyThrSerAlaProThrGlnLeuLeuArgHisProAsn 59
 QY 2799 ACATAACAGAAATAACTTGGCATCTGCCTGCACCAACCCAGGATGGAAGACATCTCC 2858
 Db 60 lIetrpHeGlyAsnIleProProProArgArgProGlnAspAsnArgAspPheSer 79
 QY 2859 CCACAACTGCCACTG---CTCACCAGGACAAGCTGCCCTTCTCTCCACCTCTCAGTC 2915
 Db 80 ProLeuHisProLeuValPhe-ProGlyHisHisSer----- 91
 QY 2916 CCCCTAGATGGATGGCTGGGAGAGGTGGAGGCTGACAGCTGAGACGTAGTGTGAGATA 2975
 Db 92 -----GlnLeuArg----- 94
 QY 2976 TGATCTAGAGGGCGGATCACCGGGATCCGGAGCATACAAGTAACATGTTTCCATGGC 3035
 Db 95 -----HisValHisGluThrGlnValG 103
 QY 3036 AACTGCTTGTCTGTTGAATTA----- 3057
 Db 103 nGlnThrCysProGlyGluLeuLysLeuSerGlyIleGluLeuProProAlaProG 123
 QY 3058 -AGACAGCAGTCAGTGTGTCATTCATGCATCAGAGGCTCTATCT---CCAGGCACAAATGTC 3113
 Db 123 mArgGlnHisSerLeuLeuProLeuHisIleThrArgProSerArgPheProHisPheHi 143
 QY 3114 CCTGTGCTCTCCTAATCCAATGGACTTCTCTCACCAGGATGATAACACCCAGAACT 3173
 Db 143 salalargArgProAspValLeuProSerLeuProAspHisGly-----ProVa 159
 QY 3174 CACTTCTCAGTCATCTCCACAG---CCGATGACTCAGAGAGCCCAACCCAGATGGGCG 3230
 Db 159 lLeuAlaGluThrLysProArgThrSerValArgGlnProArgSerThrArgGlyPr 179
 QY 3231 CTCTCTTTTCCCATCACACA-----GACTCCCTGCACAACT 3266
 Db 179 oSerPheArgProIleLeuLeuProGluValValHisAspAspProHisSe 199
 QY 3267 TTCCTGGCGTAACACTAGAGGAGTCCCAGTGCAGGATAGGCCCTTAACGTTTTGTTAAATAA 3326
 Db 199 rSerLeuArgProGlyArgSerArgSer-----Ar 209
 QY 3327 ACAGGTGCATGAAGAGGAGCTAAGGCCATTGTGATATCCACTCTCTCTTT----- 3378
 Db 209 gGlnLeuGlnProThrIleArgArgProLeuLeuAlaProAsnGlnPheAsnSerProAr 229
 QY 3379 -----CCACTTCTCTCTCATCTTTTCTCCATGTTTATGCTTCTCTCTGATCCCT 3428
 Db 229 gGlnProProProLeuSerAspAspProGlyIleLeuGlyProCys-ProLeuAlaProA 249
 QY 3429 CTTCTGCTGCACACAGACCCAGCCCTTTATTCCTCTCTCCATTTT-----CACTCC 3482
 Db 249 snSerThrArgAspProPro-ProArgProIleThrProGlyProPheAsnThrHisGly 268
 QY 3483 TTCAGCCTCTGTCCTG----- 3500
 Db 269 ValArgProLeuSerValLeuProArgAlaProProArgArgGlyLeuLeuProAsnPro 288
 QY 3501 -----AACTGCCACTGGCAACCATGGGA 3524
 Db 289 ArgArgHisArgThrSerThrGlyHisIleProSerThrThrAlaSerArgProThrGly 308
 QY 3525 CCTCAGGAC-----CAGAGACTGCTTGACTCATCTCTGGGAGGAGTAAGTTACCGGGGAC 3578
 Db 3525 CCTCAGGAC-----CAGAGACTGCTTGACTCATCTCTGGGAGGAGTAAGTTACCGGGGAC 3578

[illegible]

QY	1509	TCGCTCCTCATTTGCCACTGAGAGGTGTTTAATGTGATCCA-----AGACCTCTCT 1555
Db	209	oGluSerIysThrProLeuThrAlaThrLeuSerIysSerGlyAsnThrPheArgProG1 229
QY	1557	CTGTGAACAATTACCCCGCCAAACCACTCAG 1587
Db	229	uValHisLeuLeuProProSerGluGlu 239
RESULT 15		
T46500		
hypothetical protein DKFZp434D098.1 - human (fragment)		
C:Species: Homo sapiens (man)		
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000		
C:Accession: T46500		
R:Ottenwaelder, B.; Obermaier, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.		
submitted to the Protein Sequence Database, January 2000		
A:Reference number: Z23031		
A:Accession: T46500		
A>Status: preliminary		
A:Molecule type: mRNA		
A:Residues: 1-649 <AAA>		
A:Cross-references: EMBL:AL137564		
A:Experimental source: adult testis; clone DKFZp434D098		
C:Genetics:		
A:Note: DKFZp434D098.1		
Alignment Scores:		
Pred. No.:	1.43	Length: 649
Score:	102.50	Matches: 95
Percent Similarity:	32.70%	Conservative: 42
Best Local Similarity:	22.67%	Mismatches: 137
Query Match:	1.28%	Indels: 145
DB:	2	Gaps: 23
US-10-776-827-10 (1-4460) x T46500 (1-649)		
Qy	2650	ATTGGCCCTCTCACCCCTATATGCGCATAAACTGCCTAAGCACTCAGGCCTCCCACTCATC 2709
Db	14	MetCysProGluGlyArgTyrGly-----HisProGlyTyrProAlaLeu 28
Qy	2710	AACCCCTTTGACCAGAGAAAAGCACTCTGGTTCTCTATCCCTTGTTCACATAGAGAGT 2769
Db	29	valThrTyrSerTyr-----GlyGlyAlaValProSerTyr----- 40
Qy	2770	TTGTGATGGGGCCTCTGGCTGTGCCCTTCACATAACAGATA-----ACTGCCATCTG 2823
Db	41	-----CysProAlaTyrGlyArgValProHisSerCysGlySer 53
Qy	2824	CCTGCA-----CGAAACCCAGGGATGTGGAAGACATCTCCCCCACTGC 2868
Db	54	ProGlyGluGlyArgGlyTyrProSerProGly-----AlaHisSerProArgAlaGly 71
Qy	2869	CACTGTCTCACCAGACAAAGCTGCCCTTCTCTCCACTCTCAGTCCCCCTAGAAATGGA 2928
Db	72	SerIleSerProGlySerProProTyrProGlnSer----- 83
Qy	2929	TGGCTGGGGAGAGGTGGAGGCTGACAGCTGAGACGTAGTGTGCATATGATCTA----- 2982
Db	84	-----ArgLysLeuSerTyrGluIleProThr 92
Qy	2983	-----GGAGGGCGGATCACCGGGATCGGGACCATACAGTAACATGGTTTCCATGCCA 3036
Db	93	GluGluGlyGlyAspArgTyrProLeuProGlyHisLeuAlaSerAlaGlyProLeuAla 112
Qy	3037	ACTGCTTGCTCGTTTGAATTAAGACAGCAGTCAGTTGTCTTCATGCATGACAAAGCCCTCTA 3096
Db	113	SerAlaGluSerLeuGlu-----ProVal 120
Qy	3097	TCT-----CCAGGCACAATGTCCCTGCTGTCTCTCAATCAATGAGCTTGCTC 3144
Db	121	SerTrpArgGluGlyProSerGlyHisSerThrLeuProArgSerProArgAsp----- 138

QY	1509	TCGCTCCTCATTTGCCACTGAGAGGTGTTTAATGTGATCCA-----AGACCTCTCT 1555
Db	209	oGluSerIysThrProLeuThrAlaThrLeuSerIysSerGlyAsnThrPheArgProG1 229
QY	1557	CTGTGAACAATTACCCCGCCAAACCACTCAG 1587
Db	229	uValHisLeuLeuProProSerGluGlu 239
RESULT 15		
T46500		
hypothetical protein DKFZp434D098.1 - human (fragment)		
C:Species: Homo sapiens (man)		
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000		
C:Accession: T46500		
R:Ottenwälder, B.; Obermaier, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.		
submitted to the Protein Sequence Database, January 2000		
A:Reference number: Z23031		
A:Accession: T46500		
A>Status: preliminary		
A:Molecule type: mRNA		
A:Residues: 1-649 <AAA>		
A:Cross-references: EMBL:AL137564		
A:Experimental source: adult testis; clone DKFZp434D098		
C:Genetics:		
A:Note: DKFZp434D098.1		
Alignment Scores:		
Pred. No.:	1.43	Length: 649
Score:	102.50	Matches: 95
Percent Similarity:	32.70%	Conservative: 42
Best Local Similarity:	22.67%	Mismatches: 137
Query Match:	1.28%	Indels: 145
DB:	2	Gaps: 23
US-10-776-827-10 (1-4460) x T46500 (1-649)		
Qy	2650	ATTGGCCCTCTCACCCCTATATGCGCATAAACTGCCTAAGCACTCAGGCCTCCCACTCATC 2709
Db	14	MetCysProGluGlyArgTyrGly-----HisProGlyTyrProAlaLeu 28
Qy	2710	AACCCCTTTGACCAGAGAAAAGCACTCTGGTTCTCTATCCCTTCTCATAGACAGAGT 2769
Db	29	valThrTyrSerTyr-----GlyGlyAlaValProSerTyr----- 40
Qy	2770	TTGTGATGGGGCCTCTGGCTGTGCGCTTCACATAACAGATA-----ACTGCCATCTG 2823
Db	41	-----CysProAlaTyrGlyArgValProHisSerCysGlySer 53
Qy	2824	CCTGCA-----CGAAACCCAGGGATGTGGAAGACATCTCCCCCACTGC 2868
Db	54	ProGlyGluGlyArgGlyTyrProSerProGly-----AlaHisSerProArgAlaGly 71
Qy	2869	CACGTGCTCACCAGACAAAGCTGCCCTTCCCTCTCCACCTCTCAGTCCCCCTAGATGGA 2928
Db	72	SerIleSerProGlySerProProTyrProGlnSer----- 83
Qy	2929	TGGCTGGGGAGAGGTGGAGGCTGACAGCTGAGACGTAGTGTGCATATGATCTA----- 2982
Db	84	-----ArgLysLeuSerTyrGluIleProThr 92
Qy	2983	-----GGAGGGCGGATCACCGGGATCCGGGACCATACAGTAACATGGTTTCCATGCCA 3036
Db	93	GluGluGlyGlyAspArgTyrProLeuProGlyHisLeuAlaSerAlaGlyProLeuAla 112
Qy	3037	ACTGCTTGCTCGTTTGAATTAAGACAGCAGTCAGTTGTCTTCATGCCATGACAAAGCCCTCTA 3096
Db	113	SerAlaGluSerLeuGlu-----ProVal 120
Qy	3097	TCT-----CCAGGCACCAATGTCCCTGCTGTCTCTCAATCAATGAGCTTGCTC 3144
Db	121	SerTrpArgGluGlyProSerGlyHisSerThrLeuProArgSerProArgAsp----- 138

```
QY 3145 TCACCCAGGATGAACACCCAGAACTCCTCTCTCAGTCACCTTCCA----- 3192
Db   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
139 AlaProCysSer---AlaSerSerGluLeuSerGlyProThrProLeuHisThrSer 157
QY 3193 CAGCCGATGACTCAGAGAGACCAACCCAGAAATGGGCGCTCTCTTTCCCGATCAGAC 3252
Db   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
158 SerProValGlnGlyLysGluSerThrArgGlnAspThrArgSerProThrSerAla 177
QY 3253 TCCCTCAGCAACCTTCTCTGGCGTAAGTAGAGGATGCC-----AGTGAGGATAGGCC 3306
Db   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
178 ProThrGlnArgLeuSerProGlyGluAlaLeuProValSerGlnAlaGlyThrGly 197
QY 3307 CTAACGCTTTGTTAAATAAACAGGTGCATGAAGAGCCCTAAGGCCATTGTTGATATCC 3366
Db   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
198 LysAlaProGluLeuProSerGlySerGlyProGluProLeuAlaProSerProValSer 217
QY 3367 ACTCTCTTCTTCCACTCTCTCTCATCTTTTCTCATGTTTATGCTTCTCTGATATCC 3426
Db   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
218 ProThrPheProPro-----SerSer-Pr 225
QY 3427 CTCTCTCTGCTGCACACAGCCAGCCCGCCCTTTATTC-----TCTCCATTTTC 3477
Db   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
225 OSerAspTTPProGlnGluArgSerProGlyGlyHisSerAspGlyAlaSerProArgSe 245
QY 3478 ACTCTTCCAGCCTCTCTCCCTGAATG-----CCACTGGCAA---CCCATGGGAC 3525
Db   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
245 rProValProThrThrLeuProGlyLeuArgHisAlaPro-TripGlnGlyProArgGlyp 265
QY 3526 CTCAGGACACAGAGACTCTTGACTCATCTCTGGGAGGGTAAGTTCACGGGGGACAAAAAA 3585
Db   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
265 roProAsp----- 267
QY 3586 TGATTCCTAAAGAGAGGCTTCCTAGACACAGCAGGCTCCAGAAAGACA-----T 3636
Db   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
268 -----SerProAspGlySerProLeuThrProValLeuS 279
QY 3637 CCCTAGGCTGGACTTCTGAGCAGCTTTAGCCAGGCTCCGACGCGCAGCAGAGGAGGC 3696
Db   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
279 erGlnMetProTTPLeuValAlaSer---ProGluProProGlnSerSerProThrProA 298
QY 3697 CTTTCCCATGCT-----CCTTTCCCATTTG 3723
Db   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
298 laPheProLeuAlaAlaSerTyrAspThrAsnGlyLeuSerGlnProProLeuProGluL 318
QY 3724 CTCAAATGGATTCATGTTCTTTTCTTGGGGGAGCAGGAGGAGGAGGAGGAGGAGG 3776
Db   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
318 ysArgHisLeuPro-----GlyProGlyGlnGlnProGly 329
```

Search completed: September 21, 2004, 15:44:56
Job time : 197 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: September 21, 2004, 15:03:24 ; Search time 374.5 Seconds
(without alignments)
6729.836 Million cell updates/sec

Title: US-10-776-827-10
Perfect score: 8026
Sequence: 1 cggggcagcaaccaggagat.....aattaaagggtactactagctc 4460

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 3172214

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlpl
-Q=/cgn2 1/USPTO spool p/US10776827/runat 21092004 133239 20757/app query.fasta 1.4615
-DB=A_Geneseq 26Jan04 -QFMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10776827@cgn 1.1.651 @runat 21092004 133239 20757 -NCFU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq 29Jan04.*

1: Geneseqp1980s.*
2: Geneseqp1990s.*
3: Geneseqp2000s.*
4: Geneseqp2001s.*
5: Geneseqp2002s.*
6: Geneseqp2003as.*
7: Geneseqp2003bs.*
8: Geneseqp2004s.*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	507.5	6.3	99	6 ABO00556	Ab000556 Novel hum
2	267	3.3	47	5 AAM49810	Aam49810 Human KIA
3	249.5	3.1	49	5 AAM49809	Aam49809 Human KIA
4	170	2.1	88	4 AAG75725	Aag75725 Human col
5	159	2.0	28	4 AAO11854	Aao11854 Human pol
6	159	2.0	28	6 ABO00836	Ab000836 Polypepti
7	127.5	1.6	1213	6 ABU11838	Abu11838 Human MDD
8	127	1.6	390	4 AAU08755	Aau08755 Human ins
9	127	1.6	390	4 ABG01683	Abg01683 Novel hum
10	125.5	1.6	572	5 AAG80596	Aag80596 Human cad

11	120	1.5	899	4 ABG25235	Abg25235 Novel hum
12	118	1.5	487	6 ABM64771	Abm64771 Propionib
13	118	1.5	654	4 ABG08388	Abg08388 Novel hum
14	118	1.5	1078	6 ABP96069	Abp96069 Human pro
15	118	1.5	1251	6 ABU54581	Abu54581 Human NOV
16	114.5	1.4	806	4 ABG04954	Abg04954 Novel hum
17	113.5	1.4	443	4 AAB68080	Aab68080 Amino aci
18	113.5	1.4	451	3 AAY99382	Aay99382 Human PRO
19	113.5	1.4	451	4 AAB66131	Aab66131 Protein o
20	113.5	1.4	451	4 AAU29164	Aau29164 Human PRO
21	113.5	1.4	451	4 AAB87566	Aab87566 Human PRO
22	113.5	1.4	451	4 AAU02749	Aau02749 Novel Hum
23	113.5	1.4	451	4 AAB48069	Aab48069 Human ext
24	113.5	1.4	451	5 ABG5891	Abg5891 Human sec
25	113.5	1.4	451	5 AAU78174	Aau78174 Human cho
26	113.5	1.4	451	6 ABU58540	Abu58540 Human PRO
27	113.5	1.4	451	6 ABU88088	Abu88088 Novel hum
28	113.5	1.4	451	6 ABU84403	Abu84403 Human sec
29	113.5	1.4	451	6 ABR66277	AbR66277 Human sec
30	113.5	1.4	451	6 ABR65667	AbR65667 Human sec
31	113.5	1.4	451	6 ABU99607	Abu99607 Human sec
32	113.5	1.4	451	6 ABU82846	Abu82846 Human PRO
33	113.5	1.4	451	6 ABU89967	Abu89967 Novel hum
34	113.5	1.4	451	6 ABR68216	AbR68216 Human sec
35	113.5	1.4	451	6 ABU96269	Abu96269 Novel hum
36	113.5	1.4	451	6 ABU92700	Abu92700 Human sec
37	113.5	1.4	451	6 ABO08777	Ab008777 Human sec
38	113.5	1.4	451	6 ABO02829	Ab002829 Human sec
39	113.5	1.4	451	6 ABR74983	AbR74983 Human sec
40	113.5	1.4	451	6 ABR94745	AbR94745 Human sec
41	113.5	1.4	451	6 ABU85718	Abu85718 Human PRO
42	113.5	1.4	451	6 ABU98878	Abu98878 Novel hum
43	113.5	1.4	451	6 ABU98093	Abu98093 Novel hum
44	113.5	1.4	451	6 ABU91799	Abu91799 Novel hum
45	113.5	1.4	451	6 ABU89492	Abu89492 Human PRO

ALIGNMENTS

RESULT 1

ABO00556

ID ABO00556 standard; protein; 99 AA.

XX ABO00556;

AC ABO00556;

XX 06-AUG-2003 (first entry)

XX Novel human polypeptide #143.

DE

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

PF 13-SEP-2002; 2002WO-US029001.
XX
PR 13-SEP-2001; 2001US-0322511P.
PR 12-SEP-2002; 2002US-00243552.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Yang Y, Wang Z, Weng G, Ma Y;
XX
PP WPI; 2003-313249/30.
DR N-PSDB; ACD05633.
DR
XX
PT Novel nucleic acids and polypeptides for diagnosis, treatment of central
PT and peripheral nervous system diseases and neuropathies, such as
PT Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
PT lateral sclerosis.
XX
PS Claim 20; SEQ ID NO 479; 300pp; English.
XX
XX The present invention relates to the isolation of novel human
CC polynucleotide sequences and their encoding polypeptides. The novel
CC polypeptides exhibit activities relating to angiogenesis, cytokine, cell
CC proliferation, cell differentiation, antiinflammatory, and stem cell
CC growth factor activities. The polypeptides are involved in the
CC proliferation, differentiation and survival of pluripotent and totipotent
CC stem cells, and are useful for re-engineering damaged or diseased
CC tissues, transplantation, manufacture of bio-pharmaceuticals and
CC development of bio-sensors. The polypeptides can be used to manipulate
CC stem cells in culture to give rise to neuroepithelial cells that can be
CC used to augment or replace cells damaged by illness, autoimmune disease,
CC accidental damage or genetic disorders. The polypeptides induce the
CC proliferation of neural cells and regeneration of nerve and brain tissue
CC and are useful for the treatment of central and peripheral nervous system
CC diseases and neuropathies, such as Alzheimer's, Parkinson's disease,
CC Huntington's disease, amyotrophic lateral sclerosis (ALS). The
CC polypeptides are also involved in chemotactic or chemokinetic activity,
CC regulation of haematopoiesis and are useful for treating myeloid or
CC lymphoid cell disorders, platelet disorders such as thrombocytopaenia and
CC for regeneration of bone, cartilage, tendon, ligament and/or nerve tissue
CC growth, in tissue repair, healing of burns, incisions, ulcers, for
CC treating osteoporosis, osteoarthritis, bone degenerative disorders, and
CC periodontal disease. The polypeptides are also useful for gut protection
CC or regeneration and treatment of lung or liver fibrosis, reperfusion
CC injury in various tissues, various immune deficiencies and disorders
CC including severe combined immunodeficiency (SCID), bacterial or fungal
CC infections, autoimmune disorders (e.g. multiple sclerosis, rheumatoid
CC arthritis, diabetes mellitus, myasthenia gravis), allergic reactions and
CC conditions, such as asthma or other respiratory problems. The
CC polypeptides are involved in thrombolysis or thrombosis and are useful in
CC treatment of various coagulation disorders (including hereditary
CC disorders such as haemophilia) or to enhance coagulation and other
CC haemostatic events in treating wounds resulting from trauma, surgery or
CC other causes. The polypeptides exhibit immune stimulating or immune
CC suppressing activity, and are useful for treating autoimmune diseases or
CC cancer. They also inhibit the growth, infection or function of infectious
CC agents such as bacteria, fungi, viruses, effect biorythms or circadian
CC cycles of rhythms, fertility of male or female subjects, metabolism,
CC catabolism, and anabolism. AB000414-AB000749 represent the novel
CC polypeptides of the invention. Note: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 99 AA;

Alignment Scores:
Pred. No.: 2,18e-43 Length: 99
Score: 507.50 Matches: 96
Percent Similarity: 96.97% Conservative: 0
Best Local Similarity: 96.97% Mismatches: 0
Query Match: 6.32% Indels: 3
DB: 6 Gaps: 1

US-10-776-827-10 (1-4460) x ABO00556 (1-99)
QY 442 ATGGAGAGATCAGTGCCTCTTTCAGCTCTATCTGGGACACCATCTTGACCAACACCAA 501
Db 1 MetGluArgileSerAlaPhePheSerIleTppAspThrIleLeuThrIysHisGln 20
QY 502 GAAGGCATCTACACACCATCTGCCTCTGGGAGTCTCTCTGGGCTGCACCTCTTGGTGTATC 561
Db 21 GluGlyIleTyraThrIleCysLeuGlyValLeuLeuGlyLeuProLeuLeuValIle 40
QY 562 ATCACACTCTCTTTCATCTGTTCATCTGCTGAGCCACACGAGGAGAGGGGCCAG 621
Db 41 IleThrLeuLeuPheIleCysCysHisCysCysTppSerProGlyIysArgGlyGln 60
QY 622 CAGCCAGAG-----AAGAAAAGAGAGAGAGAGAGAGATCAAGAGACCTC 672
Db 61 GlnProGluLysAsnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 80
QY 673 TGGATCTCTGCTCAACCCCAAGCTTCTCCAGATGGAGAGACCATCATCTGCTGTT 729
Db 81 TrpIleSerAlaGlnProLysLeuLeuGlnMetGluLysArgProSerLeuProVal 99
RESULT 2
AAM49810
ID AAM49810 standard; protein; 47 AA.
AC AAM49810;
XX
DT 16-JUL-2002 (first entry)
XX
DE Human KIAA0040 neo cMNR region #2.
XX
KW MSI+; microsatellite instability tumour cell; neopeptide; cMNR; cDNR;
KW mononucleotide microsatellite; gene therapy; diagnosis; tumour; human;
KW dinucleotide microsatellite; cytostatic; immunisation.
XX
OS Homo sapiens.
XX
PN WO200204664-A2.
XX
PD 17-JAN-2002.
XX
PF 04-JUL-2001; 2001WO-DE002510.
PR 07-JUL-2000; 2000DE-01032608.
XX
PA (DOEB/) KNEBEL DOEBERITZ M.
XX
PI Knebel Doeberitz M, Bork P, Yuan YP, Gebert J, Woerner S;
PI Linnebacher M;
XX
DR WPI; 2002-171723/22.
XX
PT Mutant genes isolated from tumors showing microsatellite instability,
PT useful for diagnosis, treatment and prevention of tumors, also related
PT peptides and antibodies.
XX
PS Claim 3; Fig 2; 31pp; German.
XX
CC This invention describes novel genes isolated from MSI+ (microsatellite
CC instability) tumour cells, containing coding mononucleotide or
CC dinucleotide microsatellites (cMNR and cDNR), differing by mutations in
CC cMNR or cDNR from the corresponding genes of non-MSI+ (tumour) cells, and
CC encoding 'neopeptide'-containing gene products. The products of the
CC invention have cytostatic activity, are capable of inducing a specific
CC immune response (humoral and cellular) and are useful for gene therapy.
CC The products of the invention are used for the molecular investigation
CC and diagnosis of MSI+ tumors (or their precursors) and are useful for
CC prophylactic or therapeutic immunisation against MSI+ tumors. This
CC sequence represents the human KIAA0040 neo cMNR region described in the
CC disclosure of the invention
XX
SQ Sequence 47 AA;

CC carcinomas and cancers. AAH37196 to AAH37204 and AAH77789 represent
CC sequences used in the exemplification of the present invention. N.B.
CC Pages 666 to 682 and page 7053 of the sequence listing were missing at
CC time of publication, meaning no sequences are present for SEQ ID NO:1027
CC to 1052, 7921 and 7922
XX
SQ Sequence 88 AA;

Alignment Scores:
Pred. No.: 3.72e-08 Length: 88
Score: 170.00 Matches: 36
Percent Similarity: 97.37% Conservative: 1
Best Local Similarity: 94.74% Mismatches: 1
Query Match: 2.12% Indels: 1
DB: 4 Gaps: 0

US-10-776-827-10 (1-4460) x AAG75725 (1-88)

QY 2844 GTGGAAGACATCTCCCAACACTGCCACTGCTCACCAGGACAGCTGCCCTTCTCTCTC 2903
Db 51 MetGluAspIleSerProGlnLeuProLeuLeuThrArgThrSerCysProSerCysLeu 70

QY 2904 CACCTCTCAGTCCCTCCAGTGGCTGG-GGAGAGGTGGAGCTGACAC 2956
Db 71 HisLeuSerValProLeuGluTrpMetAlaGlyGlyGluValGluAlaAspSer 88

RESULT 5
AAO11854
ID AAO11854 standard; protein; 28 AA.
AC AAO11854;
XX
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 25746.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation.
XX
OS Homo sapiens.
XX
XX WO200164835-A2.
XX
XX 07-SEP-2001.
XX
XX 26-FEB-2001; 2001WO-US004927.
XX
XX 28-FEB-2000; 2000US-00515126.
XX
XX 18-MAY-2000; 2000US-00577409.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT;
XX
XX WPI; 2001-514838/56.
XX
XX N-PSDB; AAI91785.
XX
XX Isolated nucleic acids and polypeptides, useful for preventing diagnosing
XX and treating e.g. leukemia, inflammation and immune disorders.
XX
XX Claim 20; SEQ ID NO 25746; 1399pp + Sequence Listing; English.
XX
XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
XX the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, haematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activin/inhibin activity and may be useful in the diagnosis and/or

CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pot_sequences
XX
SQ Sequence 28 AA;

Alignment Scores:
Pred. No.: 2.68e-07 Length: 28
Score: 159.00 Matches: 28
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.98% Indels: 0
DB: 4 Gaps: 0

US-10-776-827-10 (1-4460) x AAO11854 (1-28)

QY 783 CCACACACAGACATTTTCAGAACCCCTGAAATAATGACATGTCCATGTCACAGTA 842
Db 1 ProHisThrAspIleSerGlyThrProGluIleMetHisTyrValHisValHisArgVal 20

QY 843 ACTACTCAACCCAGGACAACT 866
Db 21 ThrThrGlnProArgAsnLysPro 28

RESULT 6
ABO00836
ID ABO00836 standard; protein; 28 AA.
AC ABO00836;
XX
XX
DT 06-AUG-2003 (first entry)
XX
DE Polypeptide encoded by novel human contig #87.
XX
KW Human; angiogenesis; cytokine; cell proliferation; pluripotent;
KW cell differentiation; totipotent; stem cell; transplantation; bio-sensor;
KW neuroepithelial cell; autoimmune disease; neural cell; genetic disorder;
KW nerve; brain tissue; central nervous system disease;
KW peripheral nervous system disease; neuropathy; haematopoiesis; bone;
KW myeloid disorder; lymphoid cell disorder; platelet disorder; tendon;
KW regeneration; cartilage; tendon; ligament; nerve tissue growth;
KW tissue repair; wound healing; burn; ulcer; osteoporosis; cancer;
KW osteoarthritis; bone degenerative disorder; periodontal disease;
KW gut protection; lung fibrosis; liver fibrosis; reperfusion injury;
KW immune deficiency; infection; autoimmune disorder; allergic reaction;
KW thrombolytic; thrombosis; coagulation disorder; hereditary disorder;
KW biorhythm; circadian cycle; fertility; metabolism; catabolism; anabolism;
KW noctropic; neuroprotective; antiparkinsonian; anticonvulsant;
KW haemostatic; vulnery; antiulcer; osteopathic; antiarthritic;
KW vasotrophic; immunostimulant; antibacterial; fungicide; immunosuppressive;
KW antirheumatic; antidiabetic; antiasthmatic; cytostatic; virucide.
XX
XX Homo sapiens.
XX
XX WO2003023013-A2.
XX
XX 20-MAR-2003.
XX
XX 13-SEP-2002; 2002WO-US029001.
XX
XX 13-SEP-2001; 2001US-0322511P.
XX
XX 12-SEP-2002; 2002US-00243552.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Yang Y, Wang Z, Weng G, Ma Y;
XX
XX WPI; 2003-313249/30.
XX
XX N-PSDB; ACD05913.
XX
XX Novel nucleic acids and polypeptides for diagnosis, treatment of central
XX PT and peripheral nervous system diseases and neuropathies, such as

PT Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 PT lateral sclerosis.
 XX
 PS
 PS Example 3; SEQ ID NO 960; 300pp; English.
 XX
 CC The present invention relates to the isolation of novel human
 CC polynucleotide sequences and their encoding polypeptides. The novel
 CC polypeptides exhibit activities relating to angiogenesis, cytokine, cell
 CC proliferation, cell differentiation, anti-inflammatory, and stem cell
 CC growth factor activities. The polypeptides are involved in the
 CC proliferation, differentiation and survival of pluripotent and totipotent
 CC stem cells, and are useful for re-engineering damaged or diseased
 CC tissues, transplantation, manufacture of bio-pharmaceuticals and
 CC development of bio-sensors. The polypeptides can be used to manipulate
 CC stem cells in culture to give rise to neuroepithelial cells that can be
 CC used to augment or replace cells damaged by illness, autoimmune disease,
 CC accidental damage or genetic disorders. The polypeptides induce the
 CC proliferation of neural cells and regeneration of nerve and brain tissue
 CC and are useful for the treatment of central and peripheral nervous system
 CC diseases and neuropathies, such as Alzheimer's, Parkinson's disease,
 CC Huntington's disease, amyotrophic lateral sclerosis (ALS). The
 CC polypeptides are also involved in chemotactic or chemokinetic activity,
 CC regulation of haematopoiesis and are useful for treating myeloid or
 CC lymphoid cell disorders, platelet disorders such as thrombocytopaenia and
 CC for regeneration of bone, cartilage, tendon, ligament and/or nerve tissue
 CC growth, in tissue repair, healing of burns, incisions, ulcers, for
 CC treating osteoporosis, osteoarthritis, bone degenerative disorders, and
 CC periodontal disease. The polypeptides are also useful for gut protection
 CC or regeneration and treatment of lung or liver fibrosis, reperfusion
 CC injury in various tissues, various immune deficiencies and disorders
 CC including severe combined immunodeficiency (SCID), bacterial or fungal
 CC infections, autoimmune disorders (e.g. multiple sclerosis, rheumatoid
 CC arthritis, diabetes mellitus, myasthenia gravis), allergic reactions and
 CC conditions, such as asthma or other respiratory problems. The
 CC polypeptides are involved in thrombolysis or thrombosis and are useful in
 CC treatment of various coagulation disorders (including hereditary
 CC disorders such as haemophilia) or to enhance coagulation and other
 CC haemostatic events in treating wounds resulting from trauma, surgery or
 CC other causes. The polypeptides exhibit immune stimulating or immune
 CC suppressing activity, and are useful for treating autoimmune diseases or
 CC cancer. They also inhibit the growth, infection or function of infectious
 CC agents such as bacteria, fungi, viruses, effect biorythms or circadian
 CC cycles of rhythms, fertility of male or female subjects, metabolism,
 CC catabolism, and anabolism. ABO00750-ABO00950 represent polypeptides
 CC encoded by novel contigs assembled in the examples of the present
 CC invention. Note: The sequence data for this patent did not form part of
 CC the printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 28 AA;

Alignment Scores:
 Pred. No.: 2,68e-07 Length: 28
 Score: 159.00 Matches: 28
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.98% Indels: 0
 DB: 6 Gaps: 0

US-10-776-827-10 (1-4460) x ABO00936 (1-28)

QY 783 CCACACAGACATTTTCAGAACCCCTGAAATATGACATATGTCATGTCACAGAGTA 842
 |||||
 Db 1 ProHisThrAspIleSerGlyThrProGluIleMetHisTyrValHisValHisArgVal 20
 |||||
 QY 843 ACTACTCAACACGACGACAACT 866
 |||||
 Db 21 ThrThrGlnProArgAenLysPro 28
 |||||

RESULT 7
 ABU11838
 ID ABU11838 standard; protein; 1213 AA.
 XX

AC ABU11838;
 XX
 DT 13-FEB-2003 (first entry)
 XX
 XX Human MDDT polypeptide SEQ ID 785.
 DE
 XX MDDT: human; disease detection and treatment molecule polypeptide;
 KW anti-inflammatory; immunosuppressive; osteopathic; cytostatic; anti-HIV;
 KW haemostatic; nephrotropic; antianaemic; antipsoriatic; hepatotropic;
 KW gene therapy; protein replacement therapy; cell proliferative disorder;
 KW cancer; adenocarcinoma; leukaemia; lymphoma; melanoma; sarcoma;
 KW anaemia; Crohn's disease; acquired immunodeficiency syndrome; AIDS;
 KW Goodpasture's syndrome; inflammation; osteoporosis; thrombocytopaenia;
 KW psoriasis; hepatitis.
 KW
 XX Homo sapiens.
 OS
 XX WO200279449-A2.
 PN
 XX 10-OCT-2002.
 PD
 XX 27-MAR-2002; 2002WO-US009944.
 PF
 XX 28-MAR-2001; 2001US-0279619P.
 PR 29-MAR-2001; 2001US-0280067P.
 PR 29-MAR-2001; 2001US-0280068P.
 PR 16-MAY-2001; 2001US-0291280P.
 PR 17-MAY-2001; 2001US-0291829P.
 PR 17-MAY-2001; 2001US-0291849P.
 PR 19-JUN-2001; 2001US-0299428P.
 PR 20-JUN-2001; 2001US-0299776P.
 PR 20-JUN-2001; 2001US-030001P.
 XX
 XX (INCY-) INCYTE GENOMICS INC.
 PA
 XX Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;
 PI Daffo GE, Hillman JU, Yu JY, Tuason O, Yap PE, Ansney SR;
 PI Daugherty SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstin EH;
 PI Feralta CH, David MH, Lewis SA, Chen AU, Panzer SR, Harris B;
 PI Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;
 XX
 XX WPI: 2003-058431/05.
 DR N-PSDB; ABX34828.
 DR
 XX New purified disease detection and treatment molecule proteins and
 PT polynucleotides, useful for diagnosing, treating or preventing cancers
 PT (e.g. leukemia or sarcoma), anaemia, Crohn's disease, AIDS, osteoporosis
 PT or hepatitis.
 PT
 XX Claim 27; SEQ ID NO 785; 339pp + Sequence Listing; English.
 PS
 XX This invention describes a novel disease detection and treatment molecule
 CC polypeptide (MDDT) which has anti-inflammatory, immunosuppressive,
 CC osteopathic, cytostatic, anti-HIV, haemostatic, nephrotropic,
 CC antianaemic, antipsoriatic and hepatotropic activity. The polynucleotides
 CC and the polypeptides of the invention can be used for gene therapy, of
 CC protein replacement therapy and are useful for treating a variety of
 CC diseases or conditions. These polypeptides or polynucleotides are
 CC particularly useful for diagnosing, treating or preventing cell
 CC proliferative disorders (e.g. cancers including adenocarcinoma,
 CC leukaemia, lymphoma, melanoma, myeloma or sarcoma), anaemia, Crohn's
 CC disease, acquired immunodeficiency syndrome (AIDS), Goodpasture's
 CC syndromes, inflammation, osteoporosis, thrombocytopaenia, psoriasis or
 CC hepatitis. ABU11450-ABU11845 represent the MDDT polynucleotides encoded
 CC by ABU11450-ABU11845, described in the disclosure of the invention. NOTE:
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 CC
 XX Sequence 1213 AA;

Alignment Scores:
 Pred. No.: 0.00478 Length: 1213
 XX

QY	496	CACCAAGAGGCATCTCAACACACACCATCTGCGCTGGAGATCTCTCTGGCGCTGCCACTCTTG	555
DB	817	TyrGlnAsnGluAspTyrLeuThrIleMetIleAlaIleIleAlaGlyAlaMetValVal	836
QY	556	GTGATC-----ATCACACTCTCTTCATCTGTTGCCATTCTGC-----	594
DB	837	IleValIlePheValThrValLeuValArgCysArgHisAlaSerArgPheLysAla	856
QY	595	-----TGGAGCCCCACAGGACCAAGAGGGCCAGCAGCCCA	627
DB	857	AlaGlnArgSerLysGlnGlyAlaGluTrpMetSerProAsnGlnGluAsnLysGlnAsn	876
QY	628	GAGAAGAAAAGAGAGAAGAGAAGAGATCAAGAACCTC-----TGGATCTCT	681
DB	877	LysLysLysLysArgLysLysArgLysSerProLysSerSerLeuLeuAsnPheValThr	896
QY	682	GCTCAACCCAAAGCTTCTCCAGAT-----GGAGAAGAGACCA	711
DB	897	IleGlu--GluSerLysProAspAlaValHisGluProIleAsnGlyThrIleSerL	916
QY	718	TCACTGCTGTTTAGTTAGGACGAGAGAGAGTGTTCCTTTCTGGGGCTAAGCCCTCT	777
DB	916	euProAlaGluLeuGluGlnSerIleGlyArgPheAsp---TrpGlyProAlaPro-	934
QY	778	TCTGACCACACACAGACATTCAGAACCCCTCGAAATAATGCATATGTCCATGTCCACA	837
DB	935	-----ProThrThrPheLysProAsnSerProAspLeuAlaLys-----HisTyrLys	951
QY	838	GAGTAACCTACTCAACAGGAACAAACCTCAGACTAAGTGTCCAGAGGAGGAGTCCC	897
DB	951	erAlaSerProGlnPro-AlaPheHisLeuLysProAspThrProVal-----SerVal	968
QY	898	AGGACCAAC-----GTGGACAATCTCTGGATCTGTCTTGGCAG	936
DB	969	LysLysHisHisValIleGlnGluLeuProLeuAspAenThrPheValGlyGly-	986
QY	937	CTATGTGTCCATAGCAATGCTCTTACTGTGAGCCAGGCGATGCCTCC-----CA	987
DB	987	-----CysAspThrLeuSer-LysArgSerSerThrSerSerAspHi	1000
QY	988	CCTGTCTCTGCATACCCCATGCAAGACCAAGAACATTT-----AT	1032
DB	1000	sPheSerAlaSer-----GluCysSerSerGlnGlyGlyPheLysThrLysGlyProLe	1018
QY	1033	CCATACATCTCAATATGTTCCCAAGTGTGTGCATGCAGTCAACACACACACACAA	1092
DB	1018	uHisThrArgGlnSerGlnArgValThrPheHisLeuProAsp-----	1033
QY	1093	ATTGAGGTAGCAGGTACGTGGGCAAGTATATCTGCTCATCAATGGTTCATTGGCTATGT	1152
DB	1034	-----GlySerGlnGluSerCysSerAspSerGlyLeuGlyAspHi	1047
QY	1153	ACTTTGTGCGAGGAAGTACATATCTACAGTCACAAAAATGTCT-----	1196
DB	1047	sGluProValGlySerGlyThrLeuIleSerHis-ProLeuProLeuValGlnProGlnA	1067
QY	1197	-----CATGGGAAGCCTTGGCAGATTGAGACACATATATA-----CAATTTC	1239
DB	1067	spGluPheTyrAspGlnAlaSerProAspLysArgThrGluAlaAspGlyAsnSerAsp	1087
QY	1240	CTAACCCAGAGGCCCATACACCATCTATTCCATATAAACCACTCAGGTTACAGATGCAT	1299
DB	1087	roAsnSerAspGlyProLeuGlyProArg-----	1096
QY	1300	GCTTTCTCTATTCTAACTCTACACATAAACTTTTACTGGAAGTACTCATATAATTGCATTT	1359

Db	1097	-----GlyLeuAlaGluAlaThrGluMetCysThrGlnGluCysLeuValLeuGlyHis	1115
Qv	1360	CCAGCAACCTGCTACACTGCCCCACCCCTTGTTGTGTTTATACAGACACCAAG	1413
Db	1115	erAapAen-CysTrpMetPro---ProGlyLeuGlyProTyrGlnHisProLys	1131
RESULT 8			
AAU08755			
ID	AAU08755	standard; protein; 390 AA.	
XX	AC	AAU08755;	
XX	DT	03-JAN-2002 (first entry)	
XX	DE	Human insulin-like growth factor binding protein-like polypeptide #2.	
XX	KW	Insulin-like growth factor binding protein; IGFBP, human; cancer;	
KW	KW	female reproduction; embryo development; food supplement; gene mapping;	
KW	KW	medical imaging; autoimmune disease; nervous system disease; cytostatic;	
KW	KW	cerebrovascular disease; wound healing; gynaecological; antifertility;	
XX	KW	gene therapy; vulnery.	
XX	OS	Homo sapiens.	
XX	FN	WO200175064-A2.	
XX	PD	11-OCT-2001.	
XX	PD	30-MAR-2001; 2001WO-US010462.	
XX	PR	31-MAR-2000; 2000US-00540217.	
PR	PR	23-AUG-2000; 2000US-00649167.	
PR	PR	14-FEB-2001; 2001US-00784748.	
XX	XX	(HYSE-) HYSEQ INC.	
XX	PI	Yamazaki V, Asundi V, Dirmanac RT, Liu C, Tang YT;	
XX	PI	WPI; 2001-626426/72.	
DR	DR	N-PSDB; AAS14770.	
XX	XX	New insulin-like growth factor binding protein-like polypeptide and	
PT	PT	encoding polynucleotides, useful for treating cancer, infertility, and	
PT	PT	arthritis, and for increasing wound healing.	
XX	XX	Claim 9; Page 108; 130pp; English.	
PS	XX	The invention relates to isolated insulin-like growth factor binding	
CC	CC	protein-like (IGFBP-like) polypeptides and their associated	
CC	CC	polynucleotides. The DNA sequences can be detected by contacting a sample	
CC	CC	with nucleic acid primers that anneal to the DNA and amplifying a product	
CC	CC	comprising a portion of the sequence. Detection of the product indicates	
CC	CC	the presence of DNA. The protein sequences can be detected by contacting	
CC	CC	a sample with a compound that binds to the polypeptide to form a complex.	
CC	CC	Detection of the complex indicates the presence of the protein. The	
CC	CC	sequences of the invention are useful for treating a subject having a	
CC	CC	need to inhibit activity or expression of IGFBP-like sequences. This	
CC	CC	involves administering an antagonist of the polypeptide, a polynucleotide	
CC	CC	that inhibits the expression of the nucleotide sequence or a therapeutic	
CC	CC	amount of the polypeptide that competes for its ligand and a carrier. The	
CC	CC	sequences are useful in treatment of disorders such as cancer, or to	
CC	CC	promote female reproductive health and embryo development. They can also	
CC	CC	be used in food supplements, in medical imaging and in gene mapping. The	
CC	CC	sequences can be used in the treatment and prevention of autoimmune	
CC	CC	diseases, nervous system diseases, cerebrovascular diseases and	
CC	CC	infertility and for enhancing wound healing. This sequence represents a	
CC	CC	human IGFBP-like polypeptide	
XX	XX	Sequence 390 AA;	
SQ			
Alignment Scores:			
Pred. No.:	0.00277	Length:	390

[illegible]

RESULT 10
AAG80596
ID AAG8


```
QY 466 AGCTTATCTGGACACACCATCTTGACCAACACCAAGAGGCATCTACACACCATCTGC 525
Db 213 lInLeuAsnMetGlyAsnProAlaProGluThrGlyArgTyrArgLeuAlaHis-----P 231
QY 526 CTGGAGTCTCTCTGGGCTGCGCCACTCTTGGTCATCAT-----CACCA 567
Db 231 roAlaThrProHisAspProArgThrThrProAspHisProSerHisIleLysProHisA 251
QY 568 CTCTCTTCTACTGTGTGCGCATCTGCTGGAGCCCAAGAGAGAGAGAGAGAGAGAGAG 627
Db 251 laProThrAsnGlnProAspThrLeuThrProThrHisHisHisHisHisHisHisHis 271
QY 628 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 687
Db 271 rgAspProSerProProProProArgThr-----ArgThrProArgGln-----Ser 287
QY 688 CCCAAGCTCTCCAGATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 747
Db 287 hrHisThrThrProProHisProThrThrThrArg-----299
QY 748 AGGTGTTCTCTTCTGGGGTAAGCCTCTCTTGACCAACACAGACATT-----797
Db 300 -----HisThrAspIleHisProProT 307
QY 798 -----TCAGAACCCCTGAATATGACTATGTC-----CATGTCACA 837
Db 307 hrHisThrAsnProProProProThrArgValLeuThrArgGlnLeuHisHisHisThrAsn 327
QY 838 GAGTAAGTACTCAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 897
Db 327 rgIlelle-----ProHisArgThrHisThrThrCysArgArgAsnHisHisAsnG 345
QY 898 AGGACACACGTGGACAAATCTTGATATCTGTCTGGCAGCATGTCTGCAATAGCAATGC 957
Db 345 lN-----Meta 347
QY 958 TCCTTACTGAGACCCAGCATGCTCCACCTGCTCTGCGCATACCCACATGCAAGC 1017
Db 347 snThrThrThrProPro-----ProPro-----HisThrProGlnIleThrH 361
QY 1018 ACAAGAACATTTATCCATACATCTCAATATGTTCCCAAG---TGTGTGCACATGCACG 1074
Db 361 isArgHisHisAsnThrThrLeuArgGluGlnThrGlnLysHisArgGlnHisIleHisG 381
QY 1075 TAACACACACACACA 1091
Db 381 lyThrGlnHisHisThr 386
RESULT 12
ABM64771
ID ABM64771 standard; protein; 487 AA.
XX AC ABM64771;
XX DT 20-OCT-2003 (first entry)
XX XX Propionibacterium acnes immunogenic polypeptide #29447.
XX DE
XX KW Acne vulgaris; antiseborrheic; dermatological; antibacterial;
XX KW immunostimulant; immune response; vaccine; immunogenic.
XX XX Propionibacterium acnes.
XX OS
XX FN WO200303515-A1.
XX PD 24-APR-2003.
XX XX
XX PF 11-OCT-2002; 2002WO-US032727.
XX XX
XX PR 15-OCT-2001; 2001US-00978825.
XX XX
XX PA (CORI-) CORIXA CORP.
```

```
XX
PI Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;
PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;
PI Barth B, Vallieve-Douglas J;
XX WPI; 2003-381789/36.
XX
XX New Propionibacterium acnes polypeptides and polynucleotides encoding the
PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,
PT or for stimulating an immune response specific for a P. acnes protein.
XX
PS Claim 7; SEQ ID NO 29447; 1481bp; English.
XX
CC The invention relates to an isolated polynucleotide (ACF64435-ACF64733)
CC encoding a Propionibacterium acnes protein. The invention also relates to
CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to
CC immunogenic fragments of P. acnes polypeptides. The invention
CC additionally encompasses expression vectors and host cells comprising a
CC polynucleotide of the invention; antibodies against polypeptides of the
CC polynucleotide of the invention; a polypeptide of the invention; a
CC invention; fusion proteins comprising a polypeptide of the invention; a
CC method for stimulating an immune response specific for a P. acnes
CC polypeptide and an isolated T cell population comprising T cells prepared
CC via this method; a vaccine composition (comprising P. acnes polypeptides,
CC polynucleotides, antibodies, fusion proteins, T cell populations, or
CC antigen-presenting cells that express the polypeptide); a method and kit
CC for detecting or determining the presence or absence of P. acnes in a
CC patient; and a method for inhibiting the development of P. acnes in a
CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion
CC proteins, T cell populations or antigen-presenting cells that express the
CC polypeptides are useful for diagnosing, preventing or treating acne
CC vulgaris, or for stimulating an immune response specific for a P. acnes
CC protein. The polynucleotides can also be used as probes or primers for
CC nucleic acid hybridisation. The vaccine composition is useful for the
CC stimulation of an immune response against P. acnes, or for treating acne,
CC and the kit is useful for performing a diagnostic assay. The present
CC sequence represents a specifically claimed P. acnes polypeptide which is
CC thought to contain an immunogenic region. Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 487 AA;
Alignment Scores:
Pred. No.: 0.0275 Length: 487
Score: 118.00 Matches: 102
Percent Similarity: 33.33% Conservative: 48
Best Local Similarity: 22.67% Mismatches: 129
Query Match: 1.47% Indels: 172
DB: 6 Gaps: 25
US-10-776-827-10 (1-4460) x ABM64771 (1-487)
QY 356 TACAAGTGTGAAGTTCGATCCCTTTGAGCTCAGGACCTGCTGTGAAGCGAGAGGTT 415
Db 9 TyrArgTrpSerProArgHisSerAlaGlyThrAlaThr-----21
QY 416 CTCTGGCCCTAATCTAGCCACGACCATGGAGAGATCAGTCCCTCTTCATGCTTATCT 475
Db 22 -----ThrProTrp-----LeuProSerAlaMetArgCys 31
QY 476 GGGACACCATCTTGACCAACACCAAGAGGCATCTACAACACCATCTGCTGGAGT-- 533
Db 32 LysGlnProSerValProArgThrIleSerSerLeuSerPhePro---CysTrp-SerG 50
QY 534 -----CCTCTGGGCGCTGCC-----548
Db 50 uArgThrValGlyGlyGlyValValSerProProProThrSerGlnHisLeuAr 70
QY 549 -----ACTCTTGTGTATCAT-----CACACTCTCTCTTCTGTGTGCCA-- 587
Db 70 gHisArgHisThrLeuAsnThrGlnArgLeuThrHisArgArgThrHisGlnLeuProTh 90
```


PN WO200281498-A2.

Search completed: September 21, 2004, 15:23:55
Job time : 423.5 secs

This Page Blank (uspto)